

On the Duality of Erasures and Defects

Yongjune Kim, *Student Member, IEEE* and B. V. K. Vijaya Kumar, *Fellow, IEEE*

Abstract

In this paper, the duality of erasures and defects will be investigated by comparing the binary erasure channel (BEC) and the binary defect channel (BDC). The duality holds for channel capacities, capacity achieving schemes, minimum distances, and upper bounds on the probability of failure to retrieve the original message. Also, the binary defect and erasure channel (BDEC) will be introduced by combining the properties of the BEC and the BDC. It will be shown that the capacity of the BDEC can be achieved by the coding scheme that combines the encoding for the defects and the decoding for the erasures. This coding scheme for the BDEC has two separate redundancy parts for correcting erasures and masking defects. Thus, we will investigate the problem of redundancy allocation between these two parts.

I. INTRODUCTION

The binary erasure channel (BEC) is a very well known channel of *communication* which was introduced by Elias in 1955 [1]. Due to its simplicity, it has been a starting point to design new coding schemes and analyze the properties of codes [2], [3]. In addition, coding schemes for BEC are still being actively researched since BEC is a very good model of the Internet [4]–[6].

In BEC, as shown in Fig. 1, the channel input $X \in \{0, 1\}$ is binary and the channel output $Y = \{0, 1, \varepsilon\}$ is ternary. It is assumed that the decoder knows the locations of erased bits denoted by ε . The capacity of the BEC with erasure probability α is given by [1], [7]

$$C_{\text{BEC}} = 1 - \alpha. \quad (1)$$

Elias [1] showed that random codes of rates arbitrarily close to C_{BEC} can be decoded on the BEC with an exponentially small error probability using maximum likelihood (ML) decoding.

Parts of the material in this paper were presented in part at the IEEE International Conference on Communications, Budapest, Hungary, June 2013 and the IEEE International Symposium on Information Theory, Istanbul, Turkey, July 2013.

Y. Kim and B. V. K. Vijaya Kumar are with the Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA, 15213, USA (e-mail: yongjunekim@cmu.edu, kumar@ece.cmu.edu).

In the case of BEC, ML decoding of linear codes is equivalent to solving linear equations [1], [6].

The binary defect channel (BDC) also has a long history. The BDC was introduced to model computer memory for *storage* by Kuznetsov and Tsybakov in 1974 [8]. At that time, erasable and programmable read only memories (EPROM) and random access memories (RAM) were modeled by the BDC [8]. Recently, BDC has received renewed attention for nonvolatile memories such as flash memories and phase change memories (PCM) [9]–[11]. In addition, BDC is theoretically important since write once memories (WOM), write unidirectional memories (WUM), and some other constrained memories can be considered as special cases of the BDC [12].

As shown in Fig. 2, BDC has a ternary channel state $S \in \{0, 1, \lambda\}$ whereas the channel input X and the channel output Y are binary. The state $S = 0$ corresponds to a stuck-at 0 defect that always outputs a 0 independent of its input value, the state $S = 1$ corresponds to a stuck-at 1 defect that always outputs a 1, and the state $S = \lambda$ corresponds to a normal cell that outputs the same value as its input. The probabilities of these states are $\beta/2$, $\beta/2$ (assuming a symmetric defect probability), and $1 - \beta$, respectively [13]–[15].

It is known that the capacity is $1 - \beta$ when both the encoder and the decoder know the defect information. If the decoder is aware of the defect locations, then the defects can be regarded as erasures so that the capacity is $1 - \beta$ [13], [14]. On the other hand, Kuznetsov and Tsybakov assumed that the encoder knows the defect information such as locations and stuck-at values of defects and the decoder does not have any information of defects [8]. It was shown that the capacity is also $1 - \beta$ even if only the encoder knows the defect information [13], [14]. Thus, the capacity of the BDC is given by

$$C_{\text{BDC}} = 1 - \beta. \quad (2)$$

The capacity of the BDC can be achieved by *random binning* when only the encoder knows the defect information [13], [14]. The practical coding scheme is the *additive encoding* which masks defects by adding a carefully selected binary vector [8], [16]. Masking defects is to make a codeword whose values at the locations of defects match the stuck-at values of the defects at those locations.

Heegard proved that the additive encoding (formulated as an optimization problem) and ML

decoding can achieve the capacity of a channel that has both defects and random errors [15]. Note that the additive encoding masks defects and the ML decoding corrects random errors.

In [17], an upper bound on the probability of masking failure was derived when the additive encoding is accomplished by solving the linear equations instead of solving the optimization problem. The derived upper bound is based on the weight distribution of the underlying codes. Based on the upper bound of [17], we will show that the additive encoding can achieve C_{BDC} by using random linear codes and solving a system of linear equations. In addition, numerical results show that structured linear codes such as Bose, Chaudhuri, and Hocquenghem (BCH) codes are good choices since their performance is not far from C_{BDC} .

In Section II, the BEC and the BDC will be discussed separately. Their channel properties, capacities, capacity achieving coding schemes and upper bounds on the probability of failure will be discussed comprehensively. Afterwards, we will investigate the duality of the BEC and the BDC. Basically, an erasure ε is *neither* 0 nor 1. In contrast, a stuck-at value (i.e., defect value) is *either* 0 or 1. Also, the *decoder* corrects erasures in the BEC and the *encoder* masks defects in the BDC. Both channels have similar capacities as shown in (1) and (2). In addition, both capacities can be achieved by solving the linear equations.

However, as we will show later in this paper, the BEC and the BDC have some important differences. The linear equations for the BEC can be described by an *overdetermined* system and there is only one solution that corrects all erasures. Meanwhile, the linear equations for the BDC correspond to an *underdetermined* system which allows several solutions that mask defects. In addition, the solution of linear equations for the BEC is the estimate of message or the estimate of erased bits, whereas the solution for the BDC is the parity or the the codeword.

In addition, the minimum distance and the weight distribution of the coding scheme for the BEC are controlled by the *parity check matrix*, whereas the minimum distance and the weight distribution of the coding scheme for the BDC come from the *generator matrix*. Because of these duality properties, the upper bound on the probability of decoding failure for the BEC and the upper bound on the probability of masking failure for the BDC have interesting similarities and differences. Considering that the BEC is a channel model for digital *communication* and the BDC is a channel model for digital *storage*, the duality is meaningful.

In Section III, the binary defect and erasure channel (BDEC) will be introduced. As shown in Fig. 3, the BDEC has both erasures (with erasure probability α for a normal cell) and defects

(with defect probability β). The capacity of the BDEC is given by

$$C_{\text{BDEC}} = (1 - \alpha)(1 - \beta). \quad (3)$$

We will show that the capacity of the BDEC can be achieved by a coding scheme that combines the encoding of the BDC and the decoding of the BEC. This proposed coding scheme for BDEC has two separated redundancy parts: one for correcting erasures and the other for masking defects. In order to minimize the probability of failure of correcting erasures and masking defects, the redundancy allocation between these two redundancy parts should be optimized.

During our proof that the proposed coding scheme achieves the capacity C_{BDEC} , lower bounds on these two redundancy components (achieving the capacity) can be obtained. However, these lower bounds may not be of much help in determining the redundancy allocation when the codeword length is finite.

Thus, we will investigate redundancy allocation for the BDEC. First, the optimal redundancy allocation is obtained by simulations. Then, we will derive the upper bound on the probability of failure for a finite codeword length and obtain the estimate of the optimal redundancy allocation by minimizing this upper bound instead of the probability of failure. Same methodology has been applied for the channel that has both defects and random errors in [18] and it was shown that the estimated redundancy allocation matches the optimal one very well. From the numerical results, we will show that this method to minimize the upper bound works well for the BDEC as well as the channel of defects and random errors.

The rest of the paper is as follows. Section II discusses the duality between erasures and defects. In Section III, the BDEC will be introduced and the redundancy allocation for the BDEC will be investigated. After discussing the numerical results in Section IV, we conclude in Section V.

II. DUALITY BETWEEN ERASURES AND DEFECTS

A. Binary Erasure Channel

For the BEC, the codeword most likely to have been transmitted is the one that agrees with all of received bits that have not been erased. If there is more than one such codeword, the decoding may lead to a failure. Thus, the following simple coding scheme was proposed in [1].

Encoding: A message $\mathbf{m} \in \{0, 1\}^k$ is encoded to a corresponding codeword $\mathbf{c} \in \{0, 1\}^n$ by $\mathcal{C} = \{\mathbf{c} = G\mathbf{m} \mid \mathbf{m} \in \{0, 1\}^k\}$ where \mathcal{C} is a set of codewords and the generator matrix G is an $n \times k$ matrix over $\{0, 1\}$ such that $\text{rank}(G) = k$. Note that the code rate $R = \frac{k}{n}$.

Decoding: Let $g : \mathbf{y} \in \{0, 1, \varepsilon\}^n \rightarrow \mathcal{C}$ denote the decoding rule. If the channel output \mathbf{y} is identical to one and only one codeword on the unerased bits, the decoding succeeds. If \mathbf{y} matches completely with several codewords on the unerased bits, the decoder chooses one of them randomly. Note that there exists at least one codeword that matches with \mathbf{y} on the unerased bits [1].

We will define a random variable D as follows.

$$D = \begin{cases} 0, & \mathbf{c} \neq \hat{\mathbf{c}} \text{ (decoding failure);} \\ 1, & \mathbf{c} = \hat{\mathbf{c}} \text{ (decoding success)} \end{cases} \quad (4)$$

where $\hat{\mathbf{c}}$ is the estimated codeword produced by the decoding rule of g .

The minimum distance d_{\min} of \mathcal{C} is given by

$$d_{\min} = \min_{\substack{\mathbf{c} \neq \mathbf{0} \\ H^T \mathbf{c} = \mathbf{0}}} \|\mathbf{c}\| \quad (5)$$

where the parity check matrix H is an $n \times (n - k)$ matrix such that $H^T G = \mathbf{0}$ (superscript T denotes transpose). Also, $\|\cdot\|$ represents the Hamming weight of a vector. Due to (5), any $d_{\min} - 1$ rows of H are linearly independent. If $e < d_{\min}$, all e erasures will be successfully corrected, which will be shown in Lemma 1.

The decoding rule of g can be described by the following linear equations [1].

$$G^{\mathcal{V}} \hat{\mathbf{m}} = \mathbf{y}^{\mathcal{V}} \quad (6)$$

where $\hat{\mathbf{m}}$ is the estimate of \mathbf{m} and $\mathcal{V} = \{j_1, \dots, j_v\}$ indicates the locations of v unerased bits. We use the notation of $\mathbf{y}^{\mathcal{V}} = (y_{j_1}, \dots, y_{j_v})^T$ and $G^{\mathcal{V}} = [\mathbf{g}_{j_1}^T, \dots, \mathbf{g}_{j_v}^T]^T$ where \mathbf{g}_j is the j -th row of G . Note that $G^{\mathcal{V}}$ is a $v \times k$ matrix.

In addition, we can represent the decoding rule g by the parity check matrix H instead of the generator matrix G as follows.

$$H^T \hat{\mathbf{c}} = (H^{\mathcal{E}})^T \hat{\mathbf{c}}^{\mathcal{E}} + (H^{\mathcal{V}})^T \hat{\mathbf{c}}^{\mathcal{V}} = \mathbf{0} \quad (7)$$

where $\mathcal{E} = \{i_1, \dots, i_e\}$ indicates the locations of e erased bits such that $\mathcal{E} \cup \mathcal{V} = \{1, 2, \dots, n\}$ and $n = e + v$. Note that $\hat{\mathbf{c}}^{\mathcal{E}} = (\hat{c}_{i_1}, \dots, \hat{c}_{i_e})^T$, $\hat{\mathbf{c}}^{\mathcal{V}} = (\hat{c}_{j_1}, \dots, \hat{c}_{j_v})^T$, $H^{\mathcal{E}} = [\mathbf{h}_{i_1}^T, \dots, \mathbf{h}_{i_e}^T]^T$ and $H^{\mathcal{V}} = [\mathbf{h}_{j_1}^T, \dots, \mathbf{h}_{j_v}^T]^T$ where \mathbf{h}_i is the i -th row of H .

From the channel model of BEC, it is clear that $\hat{\mathbf{c}}^\nu = \mathbf{y}^\nu = \mathbf{c}^\nu$ and we have to estimate the erased bits of \mathbf{c} , i.e., $\hat{\mathbf{c}}^\mathcal{E}$. Thus, (7) can be represented by the following linear equations.

$$(H^\mathcal{E})^T \hat{\mathbf{c}}^\mathcal{E} = \mathbf{q} \quad (8)$$

where $\mathbf{q} = (H^\nu)^T \mathbf{c}^\nu$. $\hat{\mathbf{m}}$ can be obtained from $\hat{\mathbf{c}}$. Note that $(H^\mathcal{E})^T$ is a $(n-k) \times e$ matrix.

Because of the weak law of large numbers, we can claim that $nC_{\text{BEC}} - \epsilon \leq v = n - e \leq nC_{\text{BEC}} + \epsilon$ with high probability for sufficiently large n . Assuming that $R < C_{\text{BEC}} - \epsilon$, we can claim that (6) and (8) are *overdetermined* because of $v > k$ and $n - k > e$.

Since $\dim(\mathcal{C}) = k$, there exists exactly one solution of (6) so long as $\text{rank}(G^\nu) = k$. If $\text{rank}(G^\nu) < k$, there are several solutions, which may result in decoding failure. Similarly, there exists exactly one solution of (8) so long as $\text{rank}(H^\mathcal{E}) = e$. Otherwise, there are several solutions, which may result in decoding failure.

The following Lemma and its proof have been known in coding theory community.

Lemma 1 ([19]): The upper bound on the probability of decoding failure of the decoding rule g is given by

$$P(D = 0 \mid |\mathcal{E}| = e) \leq \frac{\sum_{w=d_{\min}}^e A_w \binom{n-w}{e-w}}{\binom{n}{e}} \quad (9)$$

where A_w is the weight distribution of \mathcal{C} . Also, \mathcal{E} represents the set of erased locations in the channel output vector \mathbf{y} and $|\mathcal{E}| = e$ is the number of erasures in \mathbf{y} .

Proof: Without loss of generality, we can assume that the all-zero codeword $\mathbf{0}$ has been transmitted and there exists a nonzero codeword \mathbf{c} of Hamming weight w such that $\Psi_w(\mathbf{c}) \subseteq \mathcal{E}$ where $\Psi_w(\mathbf{c}) = \{i \mid c_i \neq 0\}$ denotes the locations of nonzero elements of \mathbf{c} and $\mathcal{E} = \{i_1, \dots, i_e\}$ denotes the locations of e erasures. From the given decoding rule, \mathbf{y} agrees with two codewords $\mathbf{0}$ and \mathbf{c} on unerased bits, which may result in decoding failure. Meanwhile, if there is no nonzero codeword \mathbf{c} such that $\Psi_w(\mathbf{c}) \subseteq \mathcal{E}$, then \mathbf{y} agrees with only $\mathbf{0}$ on the unerased bits and the decoding succeeds.

For a nonzero \mathbf{c} such that $\Psi_w(\mathbf{c}) \subseteq \mathcal{E}$, the number of possible \mathcal{E} is $\binom{n-w}{e-w}$. Due to double counting, the number of possible \mathcal{E} which results in decoding failure will be less than or equal to $\sum_{w=d_{\min}}^e A_w \binom{n-w}{e-w}$. Since the number of all possible \mathcal{E} such that $|\mathcal{E}| = e$ is $\binom{n}{e}$, the upper bound on $P(D = 0 \mid |\mathcal{E}| = e)$ is given by (9). \blacksquare

From the upper bound in Lemma 1, it is clear that $P(D = 0 \mid |\mathcal{E}| = e) = 0$ for $e < d_{\min}$. The following Lemma shows that $P(D = 0 \mid |\mathcal{E}| = e)$ can be obtained exactly for $d_{\min} \leq e \leq$

$d_{\min} + \lfloor \frac{d_{\min}-1}{2} \rfloor$ where $\lfloor x \rfloor$ represents the largest integer not greater than x .

Lemma 2: For $e \leq d_{\min} + t$ where $t = \lfloor \frac{d_{\min}-1}{2} \rfloor$, $P(D = 0 \mid |\mathcal{E}| = e)$ is given by

$$P(D = 0 \mid |\mathcal{E}| = e) = \frac{1}{2} \cdot \frac{\sum_{w=d_{\min}}^e A_w \binom{n-w}{e-w}}{\binom{n}{e}}. \quad (10)$$

Proof: Without loss of generality, we can assume that the all-zero codeword $\mathbf{0}$ has been transmitted and suppose that there exists *only one* nonzero codeword \mathbf{c} of Hamming weight w such that $\Psi_w(\mathbf{c}) \subseteq \mathcal{E}$. Since there are only two possible candidates such as $\mathbf{0}$ and \mathbf{c} to guess the transmitted codeword, (10) is true. Thus, we need to show that there exists only one nonzero codeword \mathbf{c} such that $\Psi_w(\mathbf{c}) \subseteq \mathcal{E}$ for $d_{\min} \leq e \leq d_{\min} + t$.

Suppose that there are two nonzero codewords $\mathbf{c}_1, \mathbf{c}_2 \in \mathcal{C}$ such that $\|\mathbf{c}_1\| = w_1$ and $\|\mathbf{c}_2\| = w_2$ where $d_{\min} \leq w_1 \leq w_2$. The locations of nonzero elements of \mathbf{c}_1 and \mathbf{c}_2 are given by

$$\Psi_{w_1}(\mathbf{c}_1) = \{i_{1,1}, \dots, i_{1,w_1}\}, \quad (11)$$

$$\Psi_{w_2}(\mathbf{c}_2) = \{i_{2,1}, \dots, i_{2,w_2}\}. \quad (12)$$

Let $\Psi_\alpha = \{i_1, \dots, i_\alpha\}$ denote $\Psi_\alpha = \Psi_{w_1}(\mathbf{c}_1) \cap \Psi_{w_2}(\mathbf{c}_2)$. Then $\Psi_{w_1}(\mathbf{c}_1)$ and $\Psi_{w_2}(\mathbf{c}_2)$ are given by

$$\Psi_{w_1}(\mathbf{c}_1) = \Psi_\alpha \cup \{i'_{1,1}, \dots, i'_{1,\beta_1}\}, \quad (13)$$

$$\Psi_{w_2}(\mathbf{c}_2) = \Psi_\alpha \cup \{i'_{2,1}, \dots, i'_{2,\beta_2}\} \quad (14)$$

where i'_{1,j_1} for $j_1 \in \{1, \dots, \beta_1\}$ and i'_{2,j_2} for $j_2 \in \{1, \dots, \beta_2\}$ are the reindexed locations of nonzero elements of \mathbf{c}_1 and \mathbf{c}_2 that are mutually disjoint with Ψ_α . Note that $\{i'_{1,1}, \dots, i'_{1,\beta_1}\} \cap \{i'_{2,1}, \dots, i'_{2,\beta_2}\} = \emptyset$, $\beta_1 = w_1 - \alpha$ and $\beta_2 = w_2 - \alpha$.

Due to the property of linear codes, $\mathbf{c}_3 = \mathbf{c}_1 + \mathbf{c}_2$ is also a codeword of \mathcal{C} , i.e., $\mathbf{c}_3 \in \mathcal{C}$ and $\|\mathbf{c}_3\| = \beta_1 + \beta_2$. Also, the following conditions should hold because of the definition of d_{\min} .

$$\alpha + \beta_1 \geq d_{\min}, \quad (15)$$

$$\alpha + \beta_2 \geq d_{\min}, \quad (16)$$

$$\beta_1 + \beta_2 \geq d_{\min} \quad (17)$$

Thus, we can claim that $2(\alpha + \beta_1 + \beta_2) \geq 3d_{\min}$, which results in $\alpha + \beta_1 + \beta_2 \geq d_{\min} + \lfloor \frac{d_{\min}+1}{2} \rfloor = d_{\min} + t + 1$ since $\alpha + \beta_1 + \beta_2$ is an integer.

If there exist two codewords \mathbf{c}_1 and \mathbf{c}_2 such that $\Psi_{w_1}(\mathbf{c}_1) \subseteq \mathcal{E}$ and $\Psi_{w_2}(\mathbf{c}_2) \subseteq \mathcal{E}$ (i.e., $\Psi_{w_1}(\mathbf{c}_1) \cup \Psi_{w_2}(\mathbf{c}_2) \subseteq \mathcal{E}$), it means that $e \geq \alpha + \beta_1 + \beta_2 \geq d_{\min} + t + 1$. Thus, for $d_{\min} \leq e \leq d_{\min} + \lfloor \frac{d_{\min}-1}{2} \rfloor = d_{\min} + t$, there exists at most one nonzero codeword \mathbf{c} such that $\Psi_w(\mathbf{c}) \subseteq \mathcal{E}$. ■

Theorem 3: $P(D = 0 \mid |\mathcal{E}| = e)$ is given by

$$P(D = 0 \mid |\mathcal{E}| = e) = \begin{cases} 0 & \text{for } e < d_{\min}, \\ \frac{1}{2} \cdot \frac{\sum_{w=d_{\min}}^e A_w \binom{n-w}{e-w}}{\binom{n}{e}} & \text{for } d_{\min} \leq e \leq d_{\min} + t, \\ \leq \frac{\sum_{w=d_{\min}}^e A_w \binom{n-w}{e-w}}{\binom{n}{e}} & \text{for } e > d_{\min} + t. \end{cases} \quad (18)$$

Proof: The proof comes from the definition of d_{\min} in (5), Lemma 1 and Lemma 2. ■

Theorem 4 ([19]): The decoding rule of g is a capacity achieving scheme.

Proof: The decoding failure probability is given by

$$P(D = 0) = P(D = 0, |\mathcal{E}| \leq n(\alpha + \epsilon)) + P(D = 0, |\mathcal{E}| > n(\alpha + \epsilon)) \quad (21)$$

$$\leq \sum_{e=1}^{n(\alpha+\epsilon)} P(D = 0, |\mathcal{E}| = e) + \epsilon' \quad (22)$$

$$\leq \sum_{e=1}^{n(\alpha+\epsilon)} P(D = 0 \mid |\mathcal{E}| = e) + \epsilon' \quad (23)$$

$$\leq \sum_{e=1}^{n(\alpha+\epsilon)} \frac{\sum_{w=d_{\min}}^e A_w \binom{n-w}{e-w}}{\binom{n}{e}} + \epsilon' \quad (24)$$

$$\leq \frac{n}{2^{n-k}} \sum_{e=1}^{n(\alpha+\epsilon)} \frac{\sum_{w=d_{\min}}^e \binom{n}{w} \binom{n-w}{e-w}}{\binom{n}{e}} + \epsilon' \quad (25)$$

$$\leq \frac{n}{2^{n-k}} \sum_{e=1}^{n(\alpha+\epsilon)} \frac{\sum_{w=d_{\min}}^e \binom{e}{w} \binom{n}{e}}{\binom{n}{e}} + \epsilon' \quad (26)$$

$$\leq \frac{n}{2^{n-k}} \sum_{e=1}^{n(\alpha+\epsilon)} \sum_{w=d_{\min}}^e \binom{e}{w} + \epsilon' \quad (27)$$

$$\leq \frac{n}{2^{n-k}} \sum_{e=1}^{n(\alpha+\epsilon)} 2^e + \epsilon' \quad (28)$$

$$\leq n^2 2^{k-n} (\alpha + \epsilon) 2^{n(\alpha+\epsilon)} + \epsilon' \quad (29)$$

$$= n^2 (\alpha + \epsilon) 2^{n\{R-(1-\alpha)+\epsilon\}} + \epsilon' \quad (30)$$

where we assume that $n(\alpha + \epsilon)$ is an integer without loss of generality in (21). Also, (24) follows from (9) in Lemma 1. (25) follows from the fact that there exists an $[n, k]$ binary linear code whose weight distribution is bounded by [19]

$$A_w \leq \frac{n}{2^{n-k}} \binom{n}{w}. \quad (31)$$

Also, (26) follows from $\binom{n}{w} \binom{n-w}{e-w} = \binom{e}{w} \binom{n}{e}$.

If $R < 1 - \alpha - \epsilon = C_{\text{BEC}} - \epsilon$ and n is sufficiently large, (30) goes to zero. Thus, the decoding rule of g achieves C_{BEC} . ■

Remark 5: We can show that the decoding rule g is a capacity achieving scheme without considering the weight distribution of codes. If each element of G in (6) is selected uniformly at random from $\{0, 1\}$,

$$P(\text{rank}(G^V) < k) = \frac{2^k}{2^v} = 2^{-n(C_{\text{BEC}} - R)}. \quad (32)$$

Similarly, if each element of H in (8) is selected uniformly at random from $\{0, 1\}$,

$$P(\text{rank}(H^E) < e) = \frac{2^e}{2^{n-k}} = 2^{-n(C_{\text{BEC}} - R)}. \quad (33)$$

If $R < C_{\text{BEC}}$ and n is sufficiently large, both (32) and (33) go to zero. Thus, the decoding rule of g can achieve C_{BEC} by solving either (6) or (8).

Remark 6: The computational complexity of (6) is $\mathcal{O}(k^3)$ where $k = nR$. Also, the computational complexity of (8) is $\mathcal{O}(e^3)$ where $e = n\alpha$. Though both complexities are eventually $\mathcal{O}(n^3)$, we can choose one of them to reduce the computational complexity. If $\alpha < 0.5$ and $R > 0.5$, the computational complexity of (8) is less than that of (6).

B. Binary Defect Channel

We will use the notations of [15], [17] with slight modifications for the BDC. We define an additional variable “ λ ” (denoting the defect-free state) and the channel state $S \in \{0, 1, \lambda\}$. Let “ \circ ” denote the operator $\circ : \{0, 1\} \times \{0, 1, \lambda\} \rightarrow \{0, 1\}$ by

$$x \circ s = \begin{cases} x, & \text{if } s = \lambda; \\ s, & \text{if } s \neq \lambda. \end{cases} \quad (34)$$

An n -cell memory with defects is modeled by

$$\mathbf{y} = \mathbf{x} \circ \mathbf{s} \quad (35)$$

where $\mathbf{x} \in \{0, 1\}^n$ is the channel input vector and $\mathbf{y} \in \{0, 1\}^n$ is the channel output vector. Also, $\mathbf{s} \in \{0, 1, \lambda\}^n$ is the channel state vector which has the information of defect locations and stuck-at values. Note that \circ is the vector component-wise operator. The number of defects is equal to the number of non- λ components in \mathbf{s} . The number of errors due to defects is given by

$$\|\mathbf{x} \circ \mathbf{s} - \mathbf{x}\|. \quad (36)$$

As shown in Fig. 2,

$$P(S = s) = \begin{cases} 1 - \beta, & s = \lambda; \\ \frac{\beta}{2}, & s = 0; \\ \frac{\beta}{2}, & s = 1. \end{cases} \quad (37)$$

In [15], Heegard discussed additive encoding and defined the $[n, k, l]$ partitioned linear block code (PLBC) which consists of a pair of linear subspaces $\mathcal{C}_1 \subset \{0, 1\}^n$ and $\mathcal{C}_0 \subset \{0, 1\}^n$ of dimension k and l such that $\mathcal{C}_1 \cap \mathcal{C}_0 = \{\mathbf{0}\}$. Then the direct sum is given by

$$\mathcal{C} \triangleq \mathcal{C}_1 + \mathcal{C}_0 = \{\mathbf{c} = \mathbf{c}_1 + \mathbf{c}_0 | \mathbf{c}_1 \in \mathcal{C}_1, \mathbf{c}_0 \in \mathcal{C}_0\}. \quad (38)$$

Encoding: A message $\mathbf{m} \in \{0, 1\}^k$ is encoded to a corresponding codeword \mathbf{c} as follows.

$$\mathbf{c} = \mathbf{c}_1 + \mathbf{c}_0 = G_1 \mathbf{m} + G_0 \mathbf{d} \quad (39)$$

where $\mathbf{c}_1 = G_1 \mathbf{m}$ and $\mathbf{c}_0 = G_0 \mathbf{d}$. The generator matrix for \mathbf{c}_1 is $G_1 = [I_k \quad 0_{k,l}]^T$ where I_k is the k -dimensional identity matrix and $0_{k,l}$ is the zero matrix with size of $k \times l$. Also, the generator matrix for \mathbf{c}_0 is G_0 which is an $n \times l$ matrix. Note that $k + l = n$.

Since the channel state vector \mathbf{s} is available at the encoder, the encoder should choose $\mathbf{d} \in \{0, 1\}^l$ judiciously. The optimal parity \mathbf{d} is chosen to minimize the number of errors due to defects, i.e., $\|(\mathbf{c} \circ \mathbf{s}) - \mathbf{c}\|$.

Decoding: The decoder estimates the message \mathbf{m} as follows.

$$\hat{\mathbf{m}} = \tilde{G}_1^T \mathbf{y} = H_0^T \mathbf{y} \quad (40)$$

where $\hat{\mathbf{m}}$ is the estimate of \mathbf{m} and the channel output vector $\mathbf{y} = \mathbf{c} \circ \mathbf{s}$ is given by (35). The message inverse matrix \tilde{G}_1 is defined as an $n \times k$ matrix such that $\tilde{G}_1^T G_1 = I_k$, and $\tilde{G}_1^T G_0 = 0_{k,l}$ [15]. For the BDC, the message inverse matrix \tilde{G}_1 defined by Heegard will be the systematic parity check matrix H_0 since it satisfies two conditions for the message inverse matrix.

For convenience, we will define a random variable M as follows.

$$M = \begin{cases} 0, & \|(\mathbf{c} \circ \mathbf{s}) - \mathbf{c}\| \neq 0 \text{ (masking failure);} \\ 1, & \|(\mathbf{c} \circ \mathbf{s}) - \mathbf{c}\| = 0 \text{ (masking success)} \end{cases} \quad (41)$$

The minimum distance d_0 of an $[n, k, l]$ PLBC is given by [15], [16]

$$d_0 = \min_{\substack{\mathbf{c} \neq \mathbf{0} \\ G_0^T \mathbf{c} = \mathbf{0}}} \|\mathbf{c}\| \quad (42)$$

which means that any $d_0 - 1$ rows of G_0 are linearly independent. If $u < d_0$, all u defects will be masked and $\|(\mathbf{c} \circ \mathbf{s}) - \mathbf{c}\| = 0$ (i.e., $M = 1$), which will be shown in Lemma 8.

The encoder knows the channel state vector \mathbf{s} and tries to minimize $\|(\mathbf{c} \circ \mathbf{s}) - \mathbf{c}\|$ by choosing \mathbf{d} judiciously. The encoding of PLBC includes an implicit optimization problem which can be formulated as follows [9], [11], [15].

$$\mathbf{d}^* = \underset{\mathbf{d}}{\operatorname{argmin}} \|G_0^{\mathcal{U}} \mathbf{d} + G_1^{\mathcal{U}} \mathbf{m} - \mathbf{s}^{\mathcal{U}}\| \quad (43)$$

$$= \underset{\mathbf{d}}{\operatorname{argmin}} \|G_0^{\mathcal{U}} \mathbf{d} + \mathbf{b}^{\mathcal{U}}\| \quad (44)$$

where $\mathcal{U} = \{i_1, \dots, i_u\}$ indicates the set of locations of u defects and $\mathbf{b} = G_1 \mathbf{m} - \mathbf{s}$. Thus, $\mathbf{b}^{\mathcal{U}}$ is given by

$$\mathbf{b}^{\mathcal{U}} = G_1^{\mathcal{U}} \mathbf{m} - \mathbf{s}^{\mathcal{U}} \quad (45)$$

where $\mathbf{s}^{\mathcal{U}} = (s_{i_1}, \dots, s_{i_u})^T$, $G_0^{\mathcal{U}} = [\mathbf{g}_{0,i_1}^T, \dots, \mathbf{g}_{0,i_u}^T]^T$, and $G_1^{\mathcal{U}} = [\mathbf{g}_{1,i_1}^T, \dots, \mathbf{g}_{1,i_u}^T]^T$. Note that $\mathbf{g}_{0,i}$ and $\mathbf{g}_{1,i}$ are the i -th rows of G_0 and G_1 respectively. Also, $\|G_0^{\mathcal{U}} \mathbf{d} + \mathbf{b}^{\mathcal{U}}\|$ represents the number of errors due to defects which is equivalent to (36).

By solving the optimization problem of (44), the number of errors due to defects will be minimized. However, the computational complexity for solving (44) is exponential, which is impractical [9].

Instead of solving the impractical optimization problem, C_{BDC} can be achieved by solving the following system of linear equations [10].

$$G_0^{\mathcal{U}} \mathbf{d} = \mathbf{b}^{\mathcal{U}} \quad (46)$$

Because of the weak law of large numbers, we can claim that $nC_{\text{BDC}} - \epsilon \leq n - u \leq nC_{\text{BDC}} + \epsilon$ with high probability for sufficiently large n . For $R < C_{\text{BDC}} - \epsilon$, (46) is *underdetermined* since

$G_0^{\mathcal{U}}$ is a $u \times (n - k)$ matrix. If (46) has at least one solution, the masking succeeds since $\|G_0^{\mathcal{U}}\mathbf{d} + \mathbf{b}^{\mathcal{U}}\| = 0$.

In [17], an upper bound on the probability of masking failure was derived and numerical results showed that the upper bound is tight and the performance of partitioned Bose, Chaudhuri, Hocquenghem (PBCH) codes is not far from C_{BDC} . We will show that the additive encoding achieves C_{BDC} by using the upper bound in [17], which explains why the performance of PBCH codes is good. The PBCH code is a special class of PLBC and its generator matrices and minimum distances can be designed by a similar method such as standard BCH codes [15], [18]

First, we will present the upper bound on the probability of masking failure for u defects through the following Lemma 7, 8, 9, and Theorem 10.

Lemma 7 ([17]): The lower and upper bounds on $P(M = 0 | |\mathcal{U}| = u)$ is given by

$$\begin{aligned} \frac{1}{2} \cdot P(\text{rank}(G_0^{\mathcal{U}}) < u \mid |\mathcal{U}| = u) &\leq P(M = 0 \mid |\mathcal{U}| = u) \\ &\leq P(\text{rank}(G_0^{\mathcal{U}}) < u \mid |\mathcal{U}| = u). \end{aligned} \quad (47)$$

Proof: (46) has at least one solution if and only if

$$\text{rank}(G_0^{\mathcal{U}}) = \text{rank}(G_0^{\mathcal{U}} \mid \mathbf{b}^{\mathcal{U}}) \quad (48)$$

where $(G_0^{\mathcal{U}} \mid \mathbf{b}^{\mathcal{U}})$ is the augmented matrix.

If $\text{rank}(G_0^{\mathcal{U}}) = u$, (46) has at least one solution since (48) holds. Thus, $P(M = 0 \mid |\mathcal{U}| = u) = 0$.

If $\text{rank}(G_0^{\mathcal{U}}) = u - j$ for $1 \leq j \leq u$, the last j rows of the row reduced echelon form of $G_0^{\mathcal{U}}$ are zero vectors. In order to satisfy the condition of (48), the last j elements of the column vector $\mathbf{b}^{\mathcal{U}}$ should also be zeros. The probability that the last j elements of the column vector $\mathbf{b}^{\mathcal{U}}$ are zeros is $\frac{1}{2^j}$ since $P(S = 0 \mid S \neq \lambda) = P(S = 1 \mid S \neq \lambda) = \frac{1}{2}$. Thus, $P(M = 0 \mid |\mathcal{U}| = u)$ is given by

$$P(M = 0 \mid |\mathcal{U}| = u) = \sum_{j=1}^u \frac{2^j - 1}{2^j} P(\text{rank}(G_0^{\mathcal{U}}) = u - j \mid |\mathcal{U}| = u) \quad (49)$$

which results in (47). ■

Lemma 8 ([17]): The upper bound on $P(M = 0 | |\mathcal{U}| = u)$ is given by

$$P(M = 0 \mid |\mathcal{U}| = u) \leq \frac{\sum_{w=d_0}^u B_w \binom{n-w}{u-w}}{\binom{n}{u}} \quad (50)$$

where B_w is the weight distribution of \mathcal{C}_0^\perp (i.e., the dual code of \mathcal{C}_0).

Proof: Suppose that there exists a nonzero codeword $\mathbf{c}^\perp \in \mathcal{C}_0^\perp$ of Hamming weight w . Note that G_0 is the parity check matrix of \mathcal{C}_0^\perp . Let $\Psi_w(\mathbf{c}^\perp) = \{i \mid c_i^\perp \neq 0\}$ denote the locations of nonzero elements of \mathbf{c}^\perp and $\mathcal{U} = \{i_1, \dots, i_u\}$ denote the locations of u defects.

If $\Psi_w(\mathbf{c}^\perp) \subseteq \mathcal{U}$, $\text{rank}(G_0^\mathcal{U}) < u$. The reason is that $G_0^{\Psi_w(\mathbf{c}^\perp)}$ is a submatrix of $G_0^\mathcal{U}$ and the rows of $G_0^{\Psi_w(\mathbf{c}^\perp)}$ are linearly dependent since $G_0^T \mathbf{c}^\perp = \mathbf{0}$.

For any \mathbf{c}^\perp such that $\Psi_w(\mathbf{c}^\perp) \subseteq \mathcal{U}$, the number of possible \mathcal{U} is $\binom{n-w}{u-w}$. Due to double counting, the number of \mathcal{U} which results in $\text{rank}(G_0^\mathcal{U}) < u$ will be less than or equal to $\sum_{w=d_0}^u B_w \binom{n-w}{u-w}$. Since the number of all possible \mathcal{U} such that $|\mathcal{U}| = u$ is $\binom{n}{u}$,

$$P(\text{rank}(G_0^\mathcal{U}) < u \mid |\mathcal{U}| = u) \leq \frac{\sum_{w=d_0}^u B_w \binom{n-w}{u-w}}{\binom{n}{u}}. \quad (51)$$

By (47) and (51), the upper bound on $P(M = 0 \mid |\mathcal{U}| = u)$ is given by (50). \blacksquare

From the upper bound in Lemma 8, it is clear that $P(M = 0 \mid |\mathcal{U}| = u) = 0$ for $u < d_0$. It is worth mentioning that the upper bound on $P(M = 0 \mid |\mathcal{U}| = u)$ for the BDC is similar to the upper bound on $P(D = 0 \mid |\mathcal{E}| = e)$ for the BEC presented in Lemma 1.

Similar to Lemma 2, the following Lemma shows that $P(M = 0 \mid |\mathcal{U}| = u)$ can be obtained exactly for $d_0 \leq u \leq d_0 + \lfloor \frac{d_0-1}{2} \rfloor$.

Lemma 9 ([17]): For $u \leq d_0 + t_0$ where $t_0 = \lfloor \frac{d_0-1}{2} \rfloor$, $P(M = 0 \mid |\mathcal{U}| = u)$ is given by

$$P(M = 0 \mid |\mathcal{U}| = u) = \frac{1}{2} \cdot \frac{\sum_{w=d_0}^u B_w \binom{n-w}{u-w}}{\binom{n}{u}}. \quad (52)$$

Proof: The proof has two parts. First, we will show that

$$P(\text{rank}(G_0^\mathcal{U}) < u \mid |\mathcal{U}| = u) = \frac{\sum_{w=d_0}^u B_w \binom{n-w}{u-w}}{\binom{n}{u}} \quad (53)$$

for $u \leq d_0 + t_0$, which means that there is no double counting in (51). Second, we will prove that

$$P(\text{rank}(G_0^\mathcal{U}) < u \mid |\mathcal{U}| = u) = P(\text{rank}(G_0^\mathcal{U}) = u - 1 \mid |\mathcal{U}| = u) \quad (54)$$

for $u \leq d_0 + t_0$, which means that $P(\text{rank}(G_0^\mathcal{U}) \leq u - 2 \mid |\mathcal{U}| = u) = 0$.

Then, $P(M = 0 \mid |\mathcal{U}| = u)$ is given by

$$P(M = 0 \mid |\mathcal{U}| = u) = \frac{1}{2} \cdot P(\text{rank}(G_0^\mathcal{U}) = u - 1 \mid |\mathcal{U}| = u) \quad (55)$$

$$= \frac{1}{2} \cdot \frac{\sum_{w=d_0}^u B_w \binom{n-w}{u-w}}{\binom{n}{u}} \quad (56)$$

where (55) follows from (49) and (54). Also, (56) follows from (53).

1) Proof of (53)

Suppose that there are two nonzero codewords $\mathbf{c}_1^\perp, \mathbf{c}_2^\perp \in \mathcal{C}_0^\perp$ such that $\|\mathbf{c}_1^\perp\| = w_1$ and $\|\mathbf{c}_2^\perp\| = w_2$. Without loss of generality, we can assume that $d_0 \leq w_1 \leq w_2$. The locations of nonzero elements of \mathbf{c}_1^\perp and \mathbf{c}_2^\perp are given by

$$\Psi_{w_1}(\mathbf{c}_1^\perp) = \{i_{1,1}, \dots, i_{1,w_1}\} \quad (57)$$

$$\Psi_{w_2}(\mathbf{c}_2^\perp) = \{i_{2,1}, \dots, i_{2,w_2}\}. \quad (58)$$

Let $\Psi_\alpha = \{i_1, \dots, i_\alpha\}$ denote $\Psi_\alpha = \Psi_{w_1}(\mathbf{c}_1^\perp) \cap \Psi_{w_2}(\mathbf{c}_2^\perp)$. Then $\Psi_{w_1}(\mathbf{c}_1^\perp)$ and $\Psi_{w_2}(\mathbf{c}_2^\perp)$ are given by

$$\Psi_{w_1}(\mathbf{c}_1^\perp) = \Psi_\alpha \cup \{i'_{1,1}, \dots, i'_{1,\beta_1}\} \quad (59)$$

$$\Psi_{w_2}(\mathbf{c}_2^\perp) = \Psi_\alpha \cup \{i'_{2,1}, \dots, i'_{2,\beta_2}\} \quad (60)$$

where i'_{1,j_1} for $j_1 \in \{1, \dots, \beta_1\}$ and i'_{2,j_2} for $j_2 \in \{1, \dots, \beta_2\}$ are the reindexed locations of nonzero elements of \mathbf{c}_1^\perp and \mathbf{c}_2^\perp that are mutually disjoint with Ψ_α . Note that $\{i'_{1,1}, \dots, i'_{1,\beta_1}\} \cap \{i'_{2,1}, \dots, i'_{2,\beta_2}\} = \emptyset$, $\beta_1 = w_1 - \alpha$ and $\beta_2 = w_2 - \alpha$.

Due to the property of linear codes, $\mathbf{c}_3^\perp = \mathbf{c}_1^\perp + \mathbf{c}_2^\perp$ is also a codeword of \mathcal{C}_0^\perp , i.e., $\mathbf{c}_3^\perp \in \mathcal{C}_0^\perp$ and $\|\mathbf{c}_3^\perp\| = \beta_1 + \beta_2$. Also, the following conditions should hold because of the definition of d_0 .

$$\alpha + \beta_1 \geq d_0 \quad (61)$$

$$\alpha + \beta_2 \geq d_0 \quad (62)$$

$$\beta_1 + \beta_2 \geq d_0 \quad (63)$$

Thus, we can claim that $2(\alpha + \beta_1 + \beta_2) \geq 3d_0$, which results in $\alpha + \beta_1 + \beta_2 \geq d_0 + \lfloor \frac{d_0+1}{2} \rfloor = d_0 + t_0 + 1$ since $\alpha + \beta_1 + \beta_2$ is an integer.

For double counting in (51), there should exist at least two codewords \mathbf{c}_1^\perp and \mathbf{c}_2^\perp such that $\Psi_{w_1}(\mathbf{c}_1^\perp) \cup \Psi_{w_2}(\mathbf{c}_2^\perp) \subseteq \mathcal{U}$. It means that double counting occurs only if $u \geq \alpha + \beta_1 + \beta_2 \geq d_0 + t_0 + 1$. Thus, there is no double counting for $u \leq d_0 + t_0$. For $u \leq d_0 + t_0$, there exists at most one codeword \mathbf{c}^\perp such that $\Psi_w(\mathbf{c}^\perp) \subseteq \mathcal{U}$.

2) Proof of (54)

It is clear that $\text{rank}(G_0^\mathcal{U}) = u - 1$ if and only if there exists only one nonzero codeword \mathbf{c}^\perp such that $\Psi_w(\mathbf{c}^\perp) \subseteq \mathcal{U}$. Note that $\text{rank}(G_0^\mathcal{U}) < u - 1$ if and only if \mathcal{U} includes the locations of

nonzero elements of at least two nonzero codewords. We have already shown that there exists at most one nonzero codeword \mathbf{c}^\perp such that $\Psi_w(\mathbf{c}^\perp) \subseteq \mathcal{U}$ for $u \leq d_0 + t_0$. ■

Similar to the upper bound on $P(D = 0 \mid |\mathcal{E}| = e)$ in Theorem 3 for the BEC, we can provide the upper bound on $P(M = 0 \mid |\mathcal{U}| = u)$ for the BDC as follows.

Theorem 10 ([17]): $P(M = 0 \mid |\mathcal{U}| = u)$ is given by

$$P(M = 0 \mid |\mathcal{U}| = u) = \begin{cases} 0 & \text{for } u < d_0, \\ \frac{1}{2} \cdot \frac{\sum_{w=d_0}^u B_w \binom{n-w}{u-w}}{\binom{n}{u}} & \text{for } d_0 \leq u \leq d_0 + t_0, \\ \leq \frac{\sum_{w=d_0}^u B_w \binom{n-w}{u-w}}{\binom{n}{u}} & \text{for } u > d_0 + t_0. \end{cases} \quad (64)$$

Proof: The proof comes from the definition of d_0 in (42), Lemma 8 and Lemma 9. ■

Comparing the upper bound on $P(D = 0 \mid |\mathcal{E}| = e)$ for the BEC and $P(M = 0 \mid |\mathcal{U}| = u)$ for BDC, the duality of erasures and defects can be seen. The expressions for both upper bounds in Theorem 3 and Theorem 10 are very similar. The one difference is in the minimum distances such as d_{\min} and d_0 . For the definition of d_{\min} , H is the parity check matrix in (5). Meanwhile, G_0 is the parity check matrix in (42). The other difference comes from the weight distributions such as A_w and B_w . Note that A_w is the weight distribution of \mathcal{C} and B_w the weight distribution of \mathcal{C}_0^\perp .

The following Theorem shows that the capacity of the BDC can be achieved by an encoding scheme based on solving the linear equations (46).

Theorem 11: The encoding scheme of solving the linear equations (46) is a capacity achieving scheme.

Proof: The masking failure probability is given by

$$P(M = 0) = P(M = 0, U \leq n(\beta + \epsilon)) + P(M = 0, U > n(\beta + \epsilon)) \quad (67)$$

$$\leq \sum_{u=1}^{n(\beta+\epsilon)} P(M = 0, |\mathcal{U}| = u) + \epsilon' \quad (68)$$

$$\leq \sum_{u=1}^{n(\beta+\epsilon)} P(M = 0 \mid |\mathcal{U}| = u) + \epsilon' \quad (69)$$

$$\leq \sum_{u=1}^{n(\beta+\epsilon)} \frac{\sum_{w=d_0}^u B_w \binom{n-w}{u-w}}{\binom{n}{u}} + \epsilon' \quad (70)$$

$$\leq \frac{n}{2^{n-k}} \sum_{u=1}^{n(\beta+\epsilon)} \frac{\sum_{w=d_0}^u \binom{n}{w} \binom{n-w}{u-w}}{\binom{n}{u}} + \epsilon' \quad (71)$$

$$\leq \frac{n}{2^{n-k}} \sum_{u=1}^{n(\beta+\epsilon)} \frac{\sum_{w=d_0}^u \binom{u}{w} \binom{n}{u}}{\binom{n}{u}} + \epsilon' \quad (72)$$

$$\leq \frac{n}{2^{n-k}} \sum_{u=1}^{n(\beta+\epsilon)} \sum_{w=d_0}^u \binom{u}{w} + \epsilon' \quad (73)$$

$$\leq \frac{n}{2^{n-k}} \sum_{u=1}^{n(\beta+\epsilon)} 2^u + \epsilon' \quad (74)$$

$$\leq n^2 2^{k-n} (\beta + \epsilon) 2^{n(\beta+\epsilon)} + \epsilon' \quad (75)$$

$$= n^2 (\beta + \epsilon) 2^{n\{R-(1-\beta)+\epsilon\}} + \epsilon' \quad (76)$$

where we assume that $n(\beta + \epsilon)$ is an integer without loss of generality in (67). Also, (70) follows from Lemma 8. (71) follows from the fact that \mathcal{C}_0^\perp is an $[n, k]$ linear code and the upper bound on the weight distribution of \mathcal{C}_0^\perp can be obtained from (31). (72) follows from the fact that $\binom{n}{w} \binom{n-w}{u-w} = \binom{u}{w} \binom{n}{u}$.

If $R < 1 - \beta - \epsilon = C_{\text{BDC}} - \epsilon$ and n is sufficiently large, (76) goes to zero. Thus, the additive encoding with solving the system of linear equations of (46) achieves the channel capacity of BDC. ■

It is well known that random binning is a capacity achieving scheme for the BDC. The encoding of random binning is as follows [13], [14]: Randomly partition the 2^n sequences into 2^{nR} equal size subsets (or bins) and associate a different message with each bin. When the i -th message is to be stored, search the i -th bin for a sequence (or codeword) \mathbf{c} such that $\mathbf{c} \circ \mathbf{s} = \mathbf{c}$.

The decoding is to choose the index of the bin that the channel output vector \mathbf{y} belongs to.

The encoding of random binning can be described by the following linear equations [10].

$$H_0^T \mathbf{c} = \mathbf{m} \quad (77)$$

where \mathbf{c} will be chosen to satisfy $\mathbf{c} \circ \mathbf{s} = \mathbf{c}$. We can see that the linear equations for the encoding of random binning is equivalent to (40) which represents the decoding of additive encoding, which shows *the duality between additive encoding and random binning*.

(77) can be modified into

$$H_0^T \mathbf{c} = (H_0^{\mathcal{U}})^T \mathbf{c}^{\mathcal{U}} + (H_0^{\mathcal{W}})^T \mathbf{c}^{\mathcal{W}} \quad (78)$$

$$= \mathbf{m} \quad (79)$$

where $\mathcal{U} = \{i_1, \dots, i_u\}$ indicates the locations of stuck-at defects and $\mathcal{W} = \{j_1, \dots, j_w\}$ represents the locations of normal cells such that $\mathcal{U} \cup \mathcal{W} = \{1, 2, \dots, n\}$. Note that $\mathbf{c}^{\mathcal{U}} = (c_{i_1}, \dots, c_{i_u})^T$, $\mathbf{c}^{\mathcal{W}} = (c_{j_1}, \dots, c_{j_w})^T$, $H_0^{\mathcal{U}} = [\mathbf{h}_{0,i_1}^T, \dots, \mathbf{h}_{0,i_u}^T]^T$ and $H_0^{\mathcal{W}} = [\mathbf{h}_{0,j_1}^T, \dots, \mathbf{h}_{0,j_w}^T]^T$ where $\mathbf{h}_{0,i}$ is the i -th row of H_0 . Since $\mathbf{s}^{\mathcal{U}}$ is known to the encoder, the encoder of random binning can set $\mathbf{c}^{\mathcal{U}} = \mathbf{s}^{\mathcal{U}}$. Thus, the random binning can be described by

$$(H_0^{\mathcal{W}})^T \mathbf{c}^{\mathcal{W}} = \mathbf{m}' \quad (80)$$

where $\mathbf{m}' = \mathbf{m} - (H_0^{\mathcal{U}})^T \mathbf{s}^{\mathcal{U}}$. The solution of (80) represents the codeword elements of normal cells. Note that $(H_0^{\mathcal{W}})^T$ is a $k \times (n - u)$ matrix. Thus, (80) is also *underdetermined* for $R < C_{\text{BDC}} - \epsilon$.

Remark 12: We can show that both additive encoding and random binning are capacity achieving scheme by the same method in Remark 5. If each element of G_0 in (46) is selected uniformly at random from $\{0, 1\}$,

$$P(\text{rank}(G_0^{\mathcal{U}}) < u) = \frac{2^u}{2^{n-k}} = 2^{-n(C_{\text{BDC}} - R)}. \quad (81)$$

Similarly, if each element of H_0 in (80) is selected uniformly at random from $\{0, 1\}$,

$$P(\text{rank}(H_0^{\mathcal{W}}) < u) = \frac{2^k}{2^{n-u}} = 2^{-n(C_{\text{BDC}} - R)}. \quad (82)$$

If $R < C_{\text{BDC}}$ and n is sufficiently large, both (81) and (82) go to zero. Thus, both additive encoding and random binning achieve C_{BDC} by solving linear equations.

Remark 13: The computational complexity of additive encoding of (46) is $\mathcal{O}(u^3)$ where $u = n\beta$. Also, the computational complexity of random binning of (80) is $\mathcal{O}(k^3)$ where $k = nR$. Though both complexities are $\mathcal{O}(n^3)$, we can claim that *additive encoding is better than random binning* since β is generally very small for storage systems, i.e., $u \ll k$.

C. Duality between Erasures and Defects

We will discuss the duality of erasures and defects which is summarized in Table I. In the BEC used for *communication*, the channel input $X \in \{0, 1\}$ is binary and the channel output $Y = \{0, 1, \varepsilon\}$ is ternary where the erasure ε is *neither* 0 nor 1. In the BDC used for *storage*, the channel state $S \in \{0, 1, \lambda\}$ is ternary whereas the channel input and output are binary. The ternary channel state S informs whether the given cells are stuck-at defects or normal cells. The stuck-at value is *either* 0 or 1.

The expressions for capacities of both channels are quite similar as shown in (1) and (2). In the BEC, the *decoder* corrects erasures by using the information of locations of erasures, whereas the *encoder* masks the defects by using the information of defect locations and stuck-at values in the BDC.

The capacity achieving scheme of the BEC can be represented by the linear equations based on the generator matrix G of (6) or the linear equations based on the parity check matrix H of (8). Both linear equations are *overdetermined*. The solution of the linear equations based on G is the *estimate of message* $\hat{\mathbf{m}}$ and there should be only one $\hat{\mathbf{m}}$ for decoding success. Also, the solution of the linear equations based on H is the *estimate of erased bits* $\hat{\mathbf{c}}^\varepsilon$ which should be only one $\hat{\mathbf{c}}^\varepsilon$ for decoding success.

On the other hand, the capacity achieving scheme of the BDC can be described by *underdetermined* linear equations. The *additive encoding* can be represented by the linear equations based on the generator matrix G_0 of (46) whose solution is the *parity* \mathbf{d} . Also, the *random binning* can be represented by the linear equations based on the parity check matrix H_0 of (80) whose solution is the *codeword elements of normal cells* $\mathbf{c}^\mathcal{W}$. Unlike the coding scheme of the BEC, there can be several solutions of \mathbf{d} or $\mathbf{c}^\mathcal{W}$ that matches all stuck-at defects.

We can see the duality between erasures and defects by comparing the solution $\hat{\mathbf{m}}$ of (6) and the solution \mathbf{d} of (46), i.e., *message* and *parity*. Note that coding schemes of (6) and (46) are based on the generator matrix. In addition, we can compare the duality of *codeword elements of*

erasures and codeword elements of normal cells from (46) and (80) which are coding schemes based on the parity check matrix.

In the BEC, the minimum distance d_{\min} is defined by the *parity check matrix* H , whereas the minimum distance d_0 of additive encoding for the BDC is defined by the *generator matrix* G_0 . The upper bound on the probability of decoding failure given e erasures is dependent on the weight distribution of \mathcal{C} , whereas the upper bound on the probability of masking failure given u defects is dependent on the weight distribution of \mathcal{C}_0^\perp .

If $A_w = B_w$ and $e = u$, it is clear that the upper bound on $P(D = 0 \mid |\mathcal{E}| = e)$ is same as the upper bound on $P(M = 0 \mid |\mathcal{U}| = u)$ by Theorem 3 and Theorem 10. In particular, the following Theorem shows the equivalence of the failure probabilities (i.e., the probability of correction failure of erasures and the probability of masking failure of defects).

Theorem 14: If $A_w = B_w$ and $\alpha = \beta$, then the probability of decoding failure of the BEC is same as the probability of masking failure of the BDC.

Proof: Without loss of generality, we can assume that the all-zero codeword $\mathbf{0}$ has been transmitted through the BEC. If there is only one nonzero codeword such that $\Psi(\mathbf{c}_1) \subseteq \mathcal{E}$ where \mathcal{E} indicates the location of e erasures, the decoding success probability $P(D = 1 \mid |\mathcal{E}| = e) = \frac{1}{2}$ since the decoder chooses between $\mathbf{0}$ and \mathbf{c}_1 randomly.

If there are two nonzero codewords \mathbf{c}_1 and \mathbf{c}_2 such that $\Psi(\mathbf{c}_1) \subseteq \mathcal{E}$ and $\Psi(\mathbf{c}_2) \subseteq \mathcal{E}$, it is clear that $\Psi(\mathbf{c}_3) \subseteq \mathcal{E}$ where $\mathbf{c}_3 = \mathbf{c}_1 + \mathbf{c}_2$. Since the decoder chooses one codeword from $\{\mathbf{0}, \mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_3\}$ randomly, $P(D = 1 \mid |\mathcal{E}| = e) = \frac{1}{4}$. Similarly, if there are three codewords \mathbf{c}_i such that $\Psi(\mathbf{c}_i) \subseteq \mathcal{E}$ for $i = 1, 2, 3$ and $\mathbf{c}_1 + \mathbf{c}_2 \neq \mathbf{c}_3$, $P(D = 1 \mid |\mathcal{E}| = e) = \frac{1}{8}$ since the decoder randomly chooses a codeword among $\{\mathbf{0}, \mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_3, \mathbf{c}_1 + \mathbf{c}_2, \mathbf{c}_1 + \mathbf{c}_3, \mathbf{c}_2 + \mathbf{c}_3, \mathbf{c}_1 + \mathbf{c}_2 + \mathbf{c}_3\}$. Generalizing this observation, we can claim that

$$P(D = 1 \mid |\mathcal{E}| = e) = \frac{1}{2^j} \quad (83)$$

if $\Psi(\mathbf{c}_i) \subseteq \mathcal{E}$ for $i = 0, 1, \dots, 2^j - 1$ and $\mathbf{c}_0 = \mathbf{0}$.

It is clear there are at least j codewords $\mathbf{c}^\perp \in \mathcal{C}_0^\perp$ such that $G_0^T \mathbf{c}^\perp = \mathbf{0}$ and $\Psi(\mathbf{c}^\perp) \subseteq \mathcal{U}$ for $\text{rank}(G_0^\mathcal{U}) = u - j$. From these j codewords, we can list 2^j codewords $\mathbf{c}^\perp \in \mathcal{C}_0^\perp$ such that $G_0^T \mathbf{c}^\perp = \mathbf{0}$ and $\Psi(\mathbf{c}^\perp) \subseteq \mathcal{U}$. Since the last j elements of the column vector $\mathbf{b}^\mathcal{U}$ in (48) should be zeros for masking success, we can claim that

$$P(M = 1 \mid |\mathcal{U}| = u) = \frac{1}{2^j} \quad (84)$$

if $\Psi(\mathbf{c}_i^\perp) \subseteq \mathcal{U}$ for $i = 0, 1, \dots, 2^j - 1$. It is assumed that the distribution of each element of $\mathbf{b}^\mathcal{U}$ is uniform since $P(S = 0 \mid S \neq \lambda) = P(S = 1 \mid S \neq \lambda) = \frac{1}{2}$.

If $\alpha = \beta$, the number of erasures $|\mathcal{E}|$ and the number of defects $|\mathcal{U}|$ follow an identical binomial distribution. If $A_w = B_w$, the codeword set \mathcal{C} for the BEC and the dual codeword set \mathcal{C}_0^\perp for the BDC are also identical. Thus, we can claim that $P(D = 1 \mid |\mathcal{E}| = e) = P(M = 1 \mid |\mathcal{U}| = u)$ by (83) and (84).

Since

$$P(D = 1) = \sum P(|\mathcal{E}| = e)P(D = 1 \mid |\mathcal{E}| = e), \quad (85)$$

$$P(M = 1) = \sum P(|\mathcal{U}| = u)P(U = 1 \mid |\mathcal{U}| = u), \quad (86)$$

we can claim that $P(D = 1) = P(M = 1)$. ■

In Section IV-A, we show that numerical results confirm the duality between erasures and defects.

From channel properties, capacities, capacity achieving schemes, their upper bounds, and their failure probability, we have demonstrated the duality between erasures and defects.

III. BINARY DEFECT AND ERASURE CHANNEL

A. Binary Defect and Erasure Channel

Considering the duality between erasures and defects, we now introduce the BDEC which has both erasures and defects. As shown in Fig. 3, the probability of defects are defined by (37), and normal cells behave as the BEC with parameter α . The capacity of the BDEC was given by (3).

In order to mask defects and correct erasures, the following two cases will be considered.

- Case 1: Only the decoder has knowledge of both defects and erasures.
- Case 2: The encoder has only knowledge of defects and the decoder has only knowledge of erasures.

In case 1, the decoder can regard defects as erasures. Then, a fraction $(1 - \beta)(1 - \alpha)$ of bits are unerased. The coding scheme for the BEC can achieve the channel capacity of (3) since the BDEC is equivalent to the BEC with parameter $1 - (1 - \beta)(1 - \alpha)$.

In case 2, the encoder masks the defects by the additive encoding and the decoder corrects the erasures. The proposed coding scheme for the BDEC combines the encoding of the BDC and the decoding of the BEC.

TABLE I
DUALITY BETWEEN BEC AND BDC

	BEC	BDC
Channel property	Ternary output $Y \in \{0, 1, \varepsilon\}$	Ternary state $S \in \{0, 1, \lambda\}$
Value	Erasure ε is neither “0” nor “1”	Defect is either “0” or “1”
Capacity	$C_{\text{BEC}} = 1 - \alpha$ (1)	$C_{\text{BDC}} = 1 - \beta$ (2)
Channel information	Locations	Locations and stuck-at values
Correcting / masking	Decoder corrects erasures	Encoder masks defects
Linear equation	$G^\nu \hat{\mathbf{m}} = \mathbf{y}^\nu$ (6) $(H^\varepsilon)^T \hat{\mathbf{c}}^\varepsilon = \mathbf{q}$ (8)	$G_0^\mathcal{U} \mathbf{d} = \mathbf{b}^\mathcal{U}$ (46) $(H_0^\mathcal{W})^T \mathbf{c}^\mathcal{W} = \mathbf{m}'$ (80)
Solution of linear equation	$\hat{\mathbf{m}}$ (estimate of message) or $\hat{\mathbf{c}}^\varepsilon$ (estimate of erased bits)	\mathbf{d} (parity) or $\mathbf{c}^\mathcal{W}$ (codeword elements of normal cells)
Type of linear equation	Overdetermined	Underdetermined
Minimum distance	$d_{\min} = \min\{\ \mathbf{c}\ : H^T \mathbf{c} = \mathbf{0}, \mathbf{c} \neq \mathbf{0}\}$ If $e < d_{\min}$, e erasures are corrected.	$d_0 = \min\{\ \mathbf{c}\ : G_0^T \mathbf{c} = \mathbf{0}, \mathbf{c} \neq \mathbf{0}\}$ If $u < d_0$, u defects are masked.
Upper bound on probability of failure	Theorem 3	Theorem 10
Probability of failure	If $A_w = B_w$ and $\alpha = \beta$, then $P(D = 0) = P(M = 0)$ (Theorem 14)	

Encoding: A message $\mathbf{m} \in \{0, 1\}^k$ is encoded to a codeword $\mathbf{c} = G_1 \mathbf{m} + G_0 \mathbf{d}$. Note that G_1 is an $n \times k$ generator matrix and G_0 is an $n \times l$ generator matrix such that $n > k + l$. Two generator matrices are used to correct erasures and mask defects. First, G_1 encodes a message \mathbf{m} into $G_1 \mathbf{m}$ for correcting erasures. Next, the defects will be masked by $G_0 \mathbf{d}$. The parity for masking defects \mathbf{d} will be chosen by solving (46). The encoding can be represented by

$$\mathbf{c} = \tilde{G} \begin{bmatrix} \mathbf{m} \\ \mathbf{d} \end{bmatrix} \quad (87)$$

where $\tilde{G} = [G_1 \ G_0]$ is an $n \times (k + l)$ matrix. Note that $r = n - k - l$ is the number of parity

bits for correcting erasures and l is the number of parity bits for masking defects.

Decoding: The decoding of BDEC can be done by solving the following linear equations.

$$\tilde{G}^{\mathcal{V}} \begin{bmatrix} \hat{\mathbf{m}} \\ \hat{\mathbf{d}} \end{bmatrix} = \mathbf{y}^{\mathcal{V}} \quad (88)$$

where $\mathcal{V} = \{j_1, \dots, j_v\}$ indicates the locations of v unerased bits. We use the notation of $\mathbf{y}^{\mathcal{V}} = (y_{j_1}, \dots, y_{j_v})^T$ and $\tilde{G}^{\mathcal{V}} = [\tilde{\mathbf{g}}_{j_1}^T, \dots, \tilde{\mathbf{g}}_{j_v}^T]^T$ where \mathbf{g}_j is the j -th row of \tilde{G} . By solving (88), we can obtain the estimate of message \mathbf{m} and the estimate of parity \mathbf{d} , i.e., $\hat{\mathbf{m}}$ and $\hat{\mathbf{d}}$. Note that (88) is equivalent to (6).

Also, the decoding can be done by solving the following linear equations based on the parity check matrix \tilde{H} instead of (88).

$$(\tilde{H}^{\mathcal{E}})^T \hat{\mathbf{c}}^{\mathcal{E}} = \mathbf{q}' \quad (89)$$

where $\mathbf{q}' = (\tilde{H}^{\mathcal{V}})^T \mathbf{y}^{\mathcal{V}}$.

The weight distribution of the coding scheme is defined as a pair of sets $(A_{1,w}, B_{0,w})$. $A_{1,w}$ is the weight distribution of the $[n, k+l]$ linear block code with the generator matrix $\tilde{G} = [G_1 \ G_0]$ and the parity check matrix \tilde{H} . Also, $B_{0,w}$ is the weight distribution of the $[n, k+r]$ linear block code with parity check matrix G_0 [15]. Thus, (31) will be modified into

$$A_{1,w} \leq \frac{n}{2^{n-(k+l)}} \binom{n}{w}, \quad (90)$$

$$B_{0,w} \leq \frac{n}{2^{n-(k+r)}} \binom{n}{w}. \quad (91)$$

Also, a pair of minimum distances (d_0, d_1) are defined, where d_0 represents the minimum distance for masking defects and d_1 is the minimum distance for correcting erasures in the BDEC. d_0 is same as (42) and d_1 is given by

$$d_1 = \min_{\substack{\mathbf{m} \neq \mathbf{0} \\ \tilde{H}^T \mathbf{c} = \mathbf{0}}} \|\mathbf{c}\| \quad (92)$$

Note that d_1 is greater than or equal to the minimum distance of the $[n, k+l]$ linear block code with parity check matrix \tilde{H} , while d_0 is the minimum distance of the $[n, k+r]$ linear block code with the parity check matrix G_0 [15].

In case 2, the encoder solves the linear equations of (46) in order to determine the parity \mathbf{d} for masking defects. Also, the decoder solves the linear equations of (88) to estimate \mathbf{m} . Thus,

it is clear that this coding scheme is a combination of the coding scheme for the BEC and the coding scheme for the BDC. We will now prove that this proposed coding scheme is a capacity achieving scheme.

Theorem 15: The proposed coding scheme achieves the capacity of the BDEC. The encoding and the decoding are represented by (87) and (88), respectively.

Proof: We can see that

$$P(\hat{\mathbf{m}} \neq \mathbf{m}) = P(M = 0, D = 1) + P(M = 0, D = 0) + P(M = 1, D = 0) \quad (93)$$

$$= P(M = 0) + P(M = 1, D = 0). \quad (94)$$

First, we will derive the upper bound on $P(M = 0)$, which is similar to Theorem 11. The only difference is that $B_{0,w}$ of (91) should be used instead of B_w . Thus, (76) will be changed into

$$P(M = 0) \leq n^2(\beta + \epsilon)2^{n\{\frac{k+r}{n} - (1-\beta) + \epsilon\}} + \epsilon'. \quad (95)$$

Next, the upper bound on $P(M = 1, D = 0)$ will be derived.

$$P(M = 1, D = 0) \quad (96)$$

$$= P(M = 1, D = 0, U \leq n(\beta + \epsilon), E \leq n\{(1 - \beta)\alpha + \epsilon\}) + \epsilon' \quad (97)$$

$$= \sum_{u=0}^{n(\beta+\epsilon)} \sum_{e=0}^{n((1-\beta)\alpha+\epsilon)} P(M = 1, D = 0, |\mathcal{U}| = u, |\mathcal{E}| = e) + \epsilon' \quad (98)$$

$$= \sum_{u=0}^{n(\beta+\epsilon)} \sum_{e=0}^{n((1-\beta)\alpha+\epsilon)} P(|\mathcal{U}| = u) P(M = 1 \mid |\mathcal{U}| = u) P(|\mathcal{E}| = e \mid M = 1, |\mathcal{U}| = u) \\ \cdot P(D = 0 \mid M = 1, |\mathcal{U}| = u, |\mathcal{E}| = e) + \epsilon' \quad (99)$$

$$\leq \sum_{u=0}^{n(\beta+\epsilon)} \sum_{e=0}^{n((1-\beta)\alpha+\epsilon)} P(D = 0 \mid M = 1, |\mathcal{U}| = u, |\mathcal{E}| = e) + \epsilon' \quad (100)$$

$$\leq \sum_{u=0}^{n(\beta+\epsilon)} \sum_{e=0}^{n((1-\beta)\alpha+\epsilon)} \frac{\sum_{w=d_1}^e A_{1,w} \binom{n-u-w}{e-w}}{\binom{n-u}{e}} + \epsilon' \quad (101)$$

$$\leq \frac{n}{2^{n-(k+l)}} \sum_{u=0}^{n(\beta+\epsilon)} \sum_{e=0}^{n((1-\beta)\alpha+\epsilon)} \frac{\sum_{w=d_1}^e \binom{n}{w} \binom{n-u-w}{e-w}}{\binom{n-u}{e}} + \epsilon' \quad (102)$$

$$\leq \frac{n}{2^{n-(k+l)}} \sum_{u=0}^{n(\beta+\epsilon)} \sum_{e=0}^{n((1-\beta)\alpha+\epsilon)} \sum_{w=d_1}^e \binom{e}{w} \frac{\binom{n}{w}}{\binom{n-u}{w}} + \epsilon' \quad (103)$$

$$\leq \frac{n}{2^{n-(k+l)}} \sum_{u=0}^{n(\beta+\epsilon)} \sum_{e=0}^{n((1-\beta)\alpha+\epsilon)} \sum_{w=d_1}^e \binom{e}{w} + \epsilon' \quad (104)$$

$$\leq \frac{n}{2^{n-(k+l)}} \sum_{u=0}^{n(\beta+\epsilon)} \sum_{e=0}^{n((1-\beta)\alpha+\epsilon)} 2^e + \epsilon' \quad (105)$$

$$\leq \frac{n}{2^{n-(k+l)}} \sum_{u=0}^{n(\beta+\epsilon)} \sum_{e=0}^{n((1-\beta)\alpha+\epsilon)} 2^{n((1-\beta)\alpha+\epsilon)} + \epsilon' \quad (106)$$

$$\leq n^3 \left\{ (\beta + \epsilon) + \frac{1}{n} \right\} \left\{ ((1 - \beta)\alpha + \epsilon) + \frac{1}{n} \right\} 2^{n\{\frac{k+l}{n}-1+(1-\beta)\alpha+\epsilon\}} + \epsilon' \quad (107)$$

where we assume that $n(\beta + \epsilon)$ and $n((1 - \beta)\alpha + \epsilon)$ are integers without loss of generality. (99) follows from the chain rule. (101) follows from the modification of (9) where all the defects are successfully masked and we do not need to consider the defects. Also, $A_{1,w}$ of (90) has been used instead of A_w . (103) follows from $\binom{n-u-w}{e-w} / \binom{n-u}{e} = \binom{e}{w} / \binom{n-u}{w}$. In addition, (104) follows from $\binom{n}{w} / \binom{n-u}{w} \leq 1$ for $0 \leq u \leq n$.

By (95) and (107), we can see that the upper bound on $P(\hat{\mathbf{m}} \neq \mathbf{m})$ goes to zero if n is sufficiently large and the following two conditions hold.

$$\frac{k+r}{n} < (1-\beta) - \epsilon \quad (108)$$

$$\frac{k+l}{n} < 1 - (1-\beta)\alpha - \epsilon \quad (109)$$

From the sum of (108) and (109),

$$R = \frac{k}{n} < (1-\beta)(1-\alpha) - 2\epsilon = C_{\text{BDEC}} - 2\epsilon. \quad (110)$$

Thus, the proposed coding scheme achieves the channel capacity of BDEC. \blacksquare

B. Redundancy Allocation of BDEC

The proposed coding scheme for the BDEC requires two generator matrices, namely G_0 for masking defects and G_1 for correcting erasures, which results in two parts of redundancy. Since the number of parity bits for masking defects and for correcting erasures are l and r respectively, the total redundancy is $l+r = n-k$ and the code rate is $R = k/n$.

The fact that the redundancy can be divided into two parts leads to the problem of redundancy allocation. The objective is to find an optimal redundancy allocation between l and r in order to minimize $P(\hat{\mathbf{m}} \neq \mathbf{m})$. The problem of redundancy allocation can be formulated as follows [18].

$$\begin{aligned} (\hat{l}, \hat{r}) &= \underset{(l,r)}{\operatorname{argmin}} && P(\hat{\mathbf{m}} \neq \mathbf{m}) \\ &\text{subject to} && l+r = n-k \\ &&& 0 \leq l \leq n-k \\ &&& 0 \leq r \leq n-k \end{aligned} \quad (111)$$

Not surprisingly, the optimal redundancy allocation depends on the BDEC parameters α and β . For the BDC (i.e., $\alpha = 0$), we should allot all redundancy to masking defects and the optimal redundancy allocation will be $(l^*, r^*) = (n-k, 0)$. Meanwhile, the optimal redundancy allocation for the BEC (i.e., $\beta = 0$) will be $(l^*, r^*) = (0, n-k)$, which is same as the result of [18].

When the BDEC has both defects and erasures (i.e., $\alpha \neq 0$ and $\beta \neq 0$), it is not straightforward to obtain the optimal redundancy allocation (l^*, r^*) . Without an expression for $P(\hat{\mathbf{m}} \neq \mathbf{m})$ as

a function of (l, r) , this optimization problem cannot be solved. Unfortunately, it is difficult to obtain the exact mathematical expression for $P(\hat{\mathbf{m}} \neq \mathbf{m})$.

Alternatively, we can obtain (l^*, r^*) via Monte-Carlo simulations. However, to find (l^*, r^*) by simulations requires significant computations, especially for a low $P(\hat{\mathbf{m}} \neq \mathbf{m})$. Thus, we will consider an estimate (\hat{l}, \hat{r}) which minimizes the upper bound on $P(\hat{\mathbf{m}} \neq \mathbf{m})$ instead of $P(\hat{\mathbf{m}} \neq \mathbf{m})$.

For sufficiently large n , the upper bound on $P(\hat{\mathbf{m}} \neq \mathbf{m})$ was already derived in Theorem 15 since the upper bound is the sum of (95) and (107). From (108) and (109) in Theorem 15, the required redundancy (l, r) for achieving the capacity can be given by

$$l > n(\beta + \epsilon), \quad (112)$$

$$r > n\{(1 - \beta)\alpha + \epsilon\}. \quad (113)$$

However, these asymptotic results are not useful to choose the redundancy allocation of (l, r) for a finite length code. Thus, we will derive the upper bound on $P(\hat{\mathbf{m}} \neq \mathbf{m})$ for a finite n .

We assume that the weight distributions $A_{1,w}$ and $B_{0,w}$ can be approximated by the binomial distribution as follows.

$$A_{1,w} \cong 2^{-r} \binom{n}{w} \quad (114)$$

$$B_{0,w} \cong 2^{-l} \binom{n}{w} \quad (115)$$

which hold for random codes. In addition, the weight distribution of BCH codes can be approximated by the above binomial distribution [20]. By using (114) and (115) instead of (90) and (91), the upper bound on $P(\hat{\mathbf{m}} \neq \mathbf{m})$ for a finite n will be derived in the following Theorem.

Theorem 16: For a finite n , the upper bound on $P(\hat{\mathbf{m}} \neq \mathbf{m})$ of the BDEC is given by

$$P(\hat{\mathbf{m}} \neq \mathbf{m}) \leq 2^{-l} (1 + \beta)^n + 2^{-r} \{1 + \alpha(1 - \beta)\}^n. \quad (116)$$

Proof: The proof for the finite n is similar to the proof of Theorem 15. First, the upper

bound on $P(M = 0)$ is given by

$$P(M = 0) = \sum_{u=0}^n P(|\mathcal{U}| = u) P(M = 0 \mid |\mathcal{U}| = u) \quad (117)$$

$$\leq \sum_{u=d_0}^n \binom{n}{u} \beta^u (1 - \beta)^{n-u} \frac{\sum_{w=d_0}^u B_{0,w} \binom{n-w}{u-w}}{\binom{n}{u}} \quad (118)$$

$$= 2^{-l} \sum_{u=d_0}^n \beta^u (1 - \beta)^{n-u} \sum_{w=d_0}^u \binom{n}{w} \binom{n-w}{u-w} \quad (119)$$

$$= 2^{-l} \sum_{u=d_0}^n \beta^u (1 - \beta)^{n-u} \sum_{w=d_0}^u \binom{u}{w} \binom{n}{u} \quad (120)$$

$$\leq 2^{-l} \sum_{u=0}^n \binom{n}{u} \beta^u (1 - \beta)^{n-u} \sum_{w=0}^u \binom{u}{w} \quad (121)$$

$$= 2^{-l} \sum_{u=0}^n \binom{n}{u} (2\beta)^u (1 - \beta)^{n-u} \quad (122)$$

$$= 2^{-l} (1 + \beta)^n \quad (123)$$

where (118) follows from (50) in Lemma 8 and (119) comes from (115). Also, (120) follows from $\binom{n}{w} \binom{n-w}{u-w} = \binom{u}{w} \binom{n}{u}$.

Next, the upper bound on $P(M = 1, D = 0)$ is given by

$$P(M = 1, D = 0) = \sum_{u=0}^n \sum_{e=0}^{n-u} P(M = 1, D = 0, |\mathcal{U}| = u, |\mathcal{E}| = e) \quad (124)$$

$$\leq \sum_{u=0}^n \sum_{e=0}^{n-u} P(|\mathcal{U}| = u) P(|\mathcal{E}| = e \mid |\mathcal{U}| = u) P(D = 0 \mid M = 1, |\mathcal{U}| = u, |\mathcal{E}| = e) \quad (125)$$

$$\leq \sum_{u=0}^n \binom{n}{u} \beta^u (1 - \beta)^{n-u} \sum_{e=0}^{n-u} \binom{n-u}{e} \alpha^e (1 - \alpha)^{n-u-e} \frac{\sum_{w=d_1}^e A_{1,w} \binom{n-u-w}{e-w}}{\binom{n-u}{e}} \quad (126)$$

$$= 2^{-r} \sum_{u=0}^n \binom{n}{u} \beta^u (1 - \beta)^{n-u} \sum_{e=0}^{n-u} \binom{n-u}{e} \alpha^e (1 - \alpha)^{n-u-e} \frac{\sum_{w=d_1}^e \binom{n}{w} \binom{n-u-w}{e-w}}{\binom{n-u}{e}} \quad (127)$$

$$= 2^{-r} \sum_{u=0}^n \binom{n}{u} \beta^u (1 - \beta)^{n-u} \sum_{e=0}^{n-u} \binom{n-u}{e} \alpha^e (1 - \alpha)^{n-u-e} \sum_{w=d_1}^e \binom{e}{w} \frac{\binom{n}{w}}{\binom{n-u}{w}} \quad (128)$$

$$\leq 2^{-r} \sum_{u=0}^n \binom{n}{u} \beta^u (1 - \beta)^{n-u} \sum_{e=0}^{n-u} \binom{n-u}{e} \alpha^e (1 - \alpha)^{n-u-e} \sum_{w=d_1}^e \binom{e}{w} \quad (129)$$

$$\leq 2^{-r} \sum_{u=0}^n \binom{n}{u} \beta^u (1 - \beta)^{n-u} \sum_{e=0}^{n-u} \binom{n-u}{e} (2\alpha)^e (1 - \alpha)^{n-u-e} \quad (130)$$

$$\leq 2^{-r} \sum_{u=0}^n \binom{n}{u} \beta^u (1 - \beta)^{n-u} (1 + \alpha)^{n-u} \quad (131)$$

$$= 2^{-r} \sum_{u=0}^n \binom{n}{u} \beta^u \{(1 - \beta)(1 + \alpha)\}^{n-u} \quad (132)$$

$$= 2^{-r} \{1 + \alpha(1 - \beta)\}^n \quad (133)$$

where (125) follows from the chain rule and $P(M = 1 \mid |\mathcal{E}| = e, |\mathcal{U}| = u) \leq 1$. (126) follows from (9) in Lemma 1 and (127) comes from (114). In addition, (128) is similar to (103).

Finally, (116) is obtained from (94), (123), and (133). ■

From Theorem 16, the upper bounds on $P(\hat{\mathbf{m}} \neq \mathbf{m})$ of the BEC and the BDC for a finite n can be derived as follows.

Corollary 17: For a finite n , the upper bound on $P(\hat{\mathbf{m}} \neq \mathbf{m})$ of the BEC is given by

$$P(\hat{\mathbf{m}} \neq \mathbf{m}) \leq 2^{-r} (1 + \alpha)^n, \quad (134)$$

$$\log_2 P(\hat{\mathbf{m}} \neq \mathbf{m}) \leq n \{R - 1 + \log_2 (1 + \alpha)\}. \quad (135)$$

Proof: It is clear that $P(M = 0) = 0$ and $\beta = 0$ for the BEC. By (94) and (133), the upper bound on $P(\hat{\mathbf{m}} \neq \mathbf{m})$ of the BEC is given by

$$P(\hat{\mathbf{m}} \neq \mathbf{m}) = P(M = 1, D = 0) \leq 2^{-r} (1 + \alpha)^n.$$

Also, (135) can be obtained by taking the logarithm. ■

Corollary 18: For a finite n , the upper bound on $P(\hat{\mathbf{m}} \neq \mathbf{m})$ of the BDC is given by

$$P(\hat{\mathbf{m}} \neq \mathbf{m}) \leq 2^{-l} (1 + \beta)^n \quad (136)$$

$$\log_2 P(\hat{\mathbf{m}} \neq \mathbf{m}) \leq n \{R - 1 + \log_2 (1 + \beta)\} .. \quad (137)$$

Proof: It is clear that $P(M = 1, D = 0) = 0$ and $\alpha = 0$ for the BDC. By (94) and (123), the upper bound on $P(\hat{\mathbf{m}} \neq \mathbf{m})$ of the BDC is given by

$$P(\hat{\mathbf{m}} \neq \mathbf{m}) = P(M = 0) \leq 2^{-l} (1 + \beta)^n.$$

Also, (137) can be obtained by taking the logarithm. ■

Since (\hat{l}, \hat{r}) minimizes the upper bound on $P(\hat{\mathbf{m}} \neq \mathbf{m})$, the optimization problem in (111) is given by

$$\begin{aligned} (\hat{l}, \hat{r}) = \underset{(l, r)}{\operatorname{argmin}} \quad & 2^{-l} (1 + \beta)^n + 2^{-r} \{1 + \alpha(1 - \beta)\}^n \\ \text{subject to} \quad & l + r = n - k \\ & 0 \leq l \leq n - k \\ & 0 \leq r \leq n - k \end{aligned} \quad (138)$$

where the objective function is the upper bound on $P(\hat{\mathbf{m}} \neq \mathbf{m})$ for a finite n . This objective function is intuitively reasonable since β is the probability of defects and $\alpha(1 - \beta)$ is the probability of erasures. If $\beta \geq \alpha(1 - \beta)$, we have to allot more redundancy for masking defects, i.e., $l \geq r$. Otherwise, we should allot more redundancy for correcting erasures. For $\alpha = 0$ or $\beta = 0$, we do not need to consider the above optimization problem since the solution of the BEC or the BDC is straightforward.

If the codeword length n , the information length k and the channel parameters such as α and β are given, the solution (\hat{l}, \hat{r}) of the above optimization problem can be readily obtained. For example, we will consider $[n = 1023, k = 923, l]$ PBCH codes. All possible redundancy allocation candidates of PBCH codes are presented in Table II. Since there are only 11 redundancy

TABLE II
ALL POSSIBLE REDUNDANCY ALLOCATION CANDIDATES OF $[n = 1023, k = 923, l]$ PBCH CODES

Code	l	r	d_0	d_1	Notes
0	0	100	0	21	Only correcting erasures
1	10	90	3	19	
2	20	80	5	17	
3	30	70	7	15	
4	40	60	9	13	
5	50	50	11	11	
6	60	40	13	9	
7	70	30	15	7	
8	80	20	17	5	
9	90	10	19	3	
10	100	0	21	0	Only masking defects

allocation candidates in Table II, we can readily obtain the (\hat{l}, \hat{r}) that minimizes the objective function of (138).

In addition, the objective function is *convex* if we assume that l and r are real values. Since the optimization problem is convex, we can derive the solution (\tilde{l}, \tilde{r}) of (138) by *Karush-Kuhn-Tucker* (KKT) conditions.

$$(\tilde{l}, \tilde{r}) = \begin{cases} (0, n - k), & \text{if } 2^{-l} (1 + \beta)^n \leq 2^{-r} \{1 + \alpha(1 - \beta)\}^n; \\ (n - k, 0), & \text{if } 2^{-l} (1 + \beta)^n \geq 2^{-r} \{1 + \alpha(1 - \beta)\}^n; \\ (\check{l}, \check{r}), & \text{otherwise} \end{cases} \quad (139)$$

$$(140)$$

$$(141)$$

where (\check{l}, \check{r}) is given by

$$\check{l} = \frac{1}{2} \left\{ n \left(1 + \log_2 \left(\frac{1 + \beta}{1 + \alpha(1 - \beta)} \right) \right) - k \right\}, \quad (142)$$

$$\check{r} = \frac{1}{2} \left\{ n \left(1 - \log_2 \left(\frac{1 + \beta}{1 + \alpha(1 - \beta)} \right) \right) - k \right\}. \quad (143)$$

The details of derivation are given in Appendix. (139) and (140) are easy to see. Also, (142) and (143) are intuitively reasonable since $\tilde{l} \geq \tilde{r}$ for $\beta \geq \alpha(1 - \beta)$. If $\beta < \alpha(1 - \beta)$, $\tilde{l} < \tilde{r}$.

In Section IV-B, the numerical results show that (\hat{l}, \hat{r}) and (\tilde{l}, \tilde{r}) match (l^*, r^*) very well.

IV. NUMERICAL RESULTS

A. BEC and BDC

The numerical results for the BEC and the BDC will be presented. For the BEC, the *generator matrices* of BCH codes are used for G of (6). For the BDC, the PBCH codes are used, so the *parity check matrices* of BCH codes are used for G_0 of (46) [15]. Thus, A_w for the BEC and B_w for the BDC are same.

Fig. 4 shows the probability of decoding failure (i.e., $P(D = 0)$) and its upper bound. Also, Fig. 5 shows the probability of masking failure (i.e., $P(M = 0)$) and its upper bound. The upper bounds are given by (134) and (136). Since $\alpha = \beta = 0.1$, the upper bound for the BEC is same as the upper bound for the BDC.

Note that the slope of the upper bound on $\log_2 P(\hat{\mathbf{m}} \neq \mathbf{m})$ is n as shown in Fig. 4 and Fig. 5, which can be explained by (135) and (137). Also, the x-axis intercepts (i.e., R for $\log_2 P(\hat{\mathbf{m}} \neq \mathbf{m}) = 0$) are $1 - \log_2(1 + \alpha)$ and $1 - \log_2(1 + \beta)$ for each channel.

Fig. 6 compares the probability of decoding failure of the BEC and the probability of masking failure of the BDC. By Fig. 6, we can see that $P(D = 0) = P(M = 0)$ if $A_w = B_w$ and $\alpha = \beta$, which confirms the duality in Theorem 14.

B. Redundancy Allocation for BDEC

In order to discuss the redundancy allocation for BDEC, we will consider multiple BDECs in Table III whose capacities are $C_{\text{BDEC}} = 0.95$. For these channels, we apply $[n = 1023, k = 923, l]$ PBCH codes whose all possible redundancy allocation candidates are presented in Table II.

Fig. 7 shows the simulation results for the channels of Table II. The simulation results of channel 1 (BEC) and channel 7 (BDC) are incomplete due to their impractical computational complexities. However, it should be obvious that the optimal redundancy allocation for channel 1 (BEC) will be $(l^*, r^*) = (0, 100)$. The more defects a channel has, the larger l is expected to be for the optimal redundancy allocation. Eventually, the optimal redundancy allocation for channel 7 (BDC) will be $(l^*, r^*) = (100, 0)$. The optimal l^* for all channels of Table III can be

TABLE III
BDEC WITH THE SAME $C_{\text{BDEC}} = 0.95$

Channel	α	β	Notes
1	0.0500	0	BEC
2	0.0404	0.0100	
3	0.0306	0.0200	
4	0.0253	0.0253	
5	0.0200	0.0306	
6	0.0100	0.0404	
7	0	0.0500	BDC

obtained from Fig. 7, which are presented in the second column of Table IV. The optimal r^* can be obtained by $r^* = n - k - l^*$ [18].

To find the optimal redundancy allocation (l^*, r^*) by simulation requires significant computations. Therefore, we will try to estimate the redundancy allocation from (138) instead of the simulation for estimating the optimal redundancy allocation.

First, we can readily obtain the (\hat{l}, \hat{r}) that minimizes the objective function of (138) for each channel since there only 11 redundancy allocation candidates in Table II. The estimate \hat{l} for all channels can be obtained from Fig. 8. The estimate \hat{l} for all channels are presented in the third column of Table IV. Note that $\hat{r} = n - k - \hat{l}$. Table IV shows that the estimate (\hat{l}, \hat{r}) matches the optimal redundancy allocation (l^*, r^*) very well.

Next, (\tilde{l}, \tilde{r}) can be calculated by (139)~(143) assuming that l and r are real values. The solution \tilde{l} are presented in the last column of Table IV. Table IV shows that the optimal l^* is the nearest one from \tilde{l} considering the possible redundancy allocation candidates in Table II

V. CONCLUSIONS

The duality of erasures and defects was revealed. The erasures are corrected by the decoder and the defects are masked by the encoder. The duality holds for channel capacities, capacity achieving schemes, minimum distances, upper bounds on probabilities of failure, and probabilities

TABLE IV
OPTIMAL REDUNDANCY ALLOCATION l^* AND ITS ESTIMATE \hat{l} AND \tilde{l}

Channel	l^*	\hat{l}	\tilde{l}
1	0	0	0
2	30	30	28.4
3	40	40	42.8
4	50	50	50.5
5	60	60	58.1
6	70	70	72.2
7	100	100	100

of failure. By using the upper bounds on the probability of failures, it was proved that the capacities of the BEC and the BDC can be achieved by solving overdetermined linear equations and underdetermined linear equations, respectively.

Also, the BDEC was introduced, which has both erasures and defects. The capacity of the BDEC can be achieved by the coding scheme that combines the coding schemes of the BEC and the BDC.

In addition, we investigated the redundancy allocation for the BDEC. The optimal redundancy allocation was obtained by simulations. In order to reduce the computation complexity, we proposed two methods to estimate the optimal redundancy allocation based on the upper bound on failure probability. The numerical results showed that the estimates of redundancy allocation match the optimal redundancy allocation well.

APPENDIX
DERIVATION OF (\tilde{l}, \tilde{r})

Assume that l and r are real values. Since the objective function is convex and other constraints are linear, the optimization problem of (138) is convex. The Lagrangian L is given by

$$\begin{aligned} L(l, r, \lambda_1, \lambda_2, \lambda_3, \lambda_4, \nu) = & 2^{-l} (1 + \beta)^n + 2^{-r} \{1 + \alpha (1 - \beta)\}^n \\ & + \lambda_1(-l) + \lambda_2(-r) + \lambda_3 \{l - (n - k)\} + \lambda_4 \{r - (n - k)\} \\ & + \nu \{l + r - (n - k)\} \end{aligned} \quad (144)$$

where λ_i for $i = 1, 2, 3, 4$ are the Lagrange multipliers associated with the inequality constraints and ν is the Lagrange multiplier with the equality constraint [21].

The KKT conditions are as follows.

$$\nabla L = 0 \quad (145)$$

$$-l \leq 0 \quad (146)$$

$$-r \leq 0 \quad (147)$$

$$l - (n - k) \leq 0 \quad (148)$$

$$r - (n - k) \leq 0 \quad (149)$$

$$l + r - (n - k) = 0 \quad (150)$$

$$\lambda_i \geq 0, \quad i = 1, \dots, 4 \quad (151)$$

$$\lambda_1 l = 0 \quad (152)$$

$$\lambda_2 r = 0 \quad (153)$$

$$\lambda_3 \{l - (n - k)\} = 0 \quad (154)$$

$$\lambda_4 \{r - (n - k)\} = 0 \quad (155)$$

where (145) is given by

$$\nabla L = \begin{bmatrix} -\ln 2 \cdot 2^{-l} (1 + \beta)^n \\ -\ln 2 \cdot 2^{-r} \{1 + \alpha (1 - \beta)\}^n \end{bmatrix} + \begin{bmatrix} -\lambda_1 + \lambda_3 + \nu \\ -\lambda_2 + \lambda_4 + \nu \end{bmatrix} = 0. \quad (156)$$

We will consider the following three conditions:

- $l = 0, r = n - k$

- $l = n - k, r = 0$
- $0 < l < n - k, 0 < r < n - k$
- 1) $l = 0, r = n - k$

Due to *complementary slackness*, it is clear that $\lambda_2 = \lambda_3 = 0$. Thus, (156) will be as follows.

$$-\ln 2 \cdot 2^{-l} (1 + \beta)^n - \lambda_1 + \nu = 0 \quad (157)$$

$$-\ln 2 \cdot 2^{-r} \{1 + \alpha(1 - \beta)\}^n + \lambda_4 + \nu = 0 \quad (158)$$

Since $\lambda_1 + \lambda_4 \geq 0$,

$$2^{-l} (1 + \beta)^n \leq 2^{-r} \{1 + \alpha(1 - \beta)\}^n. \quad (159)$$

which results in (139). It reveals that we have to allot all redundancy for correcting erasures if (159) is true.

- 2) $l = n - k, r = n - k$

Due to *complementary slackness*, it is clear that $\lambda_1 = \lambda_4 = 0$. Thus, (156) will be as follows.

$$-\ln 2 \cdot 2^{-l} (1 + \beta)^n + \lambda_3 + \nu = 0 \quad (160)$$

$$-\ln 2 \cdot 2^{-r} \{1 + \alpha(1 - \beta)\}^n - \lambda_2 + \nu = 0 \quad (161)$$

Since $\lambda_2 + \lambda_3 \geq 0$,

$$2^{-l} (1 + \beta)^n \geq 2^{-r} \{1 + \alpha(1 - \beta)\}^n. \quad (162)$$

which results in (140). It reveals that we have to allot all redundancy for masking defects if (162) is true.

- 3) $0 < l < n - k, 0 < r < n - k$

Due to *complementary slackness*, it is clear that $\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = 0$. Thus, (156) will be as follows.

$$-\ln 2 \cdot 2^{-l} (1 + \beta)^n + \nu = 0 \quad (163)$$

$$-\ln 2 \cdot 2^{-r} \{1 + \alpha(1 - \beta)\}^n + \nu = 0 \quad (164)$$

By (163) and (164),

$$2^{-l} (1 + \beta)^n = 2^{-r} \{1 + \alpha(1 - \beta)\}^n. \quad (165)$$

By (150) and (165), (142) and (143) can be obtained.

REFERENCES

- [1] P. Elias, "Coding for two noisy channels," in *Proc. 3rd London Symp. Inf. Theory*, London, U.K., 1955, pp. 61–76.
- [2] E. Arikan, "Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 7, pp. 3051–3073, Jul. 2009.
- [3] C. Di, D. Proietti, I. E. Telatar, T. J. Richardson, and R. L. Urbanke, "Finite-length analysis of low-density parity-check codes on the binary erasure channel," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1570–1579, Jun. 2002.
- [4] M. G. Luby, M. Mitzenmacher, M. A. Shokrollahi, and D. A. Spielman, "Efficient erasure correcting codes," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 569–584, Feb. 2001.
- [5] M. Luby, "LT codes," in *Proc. 43rd Annu. IEEE Symp. Foundations of Computer Science (FOCS)*, Vancouver, BC, Canada, Nov. 2002, pp. 271–280.
- [6] A. Shokrollahi, "Raptor codes," *IEEE Trans. Inf. Theory*, vol. 52, no. 6, pp. 2551–2567, Jun. 2006.
- [7] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Hoboken, NJ: Wiley-Interscience, 2006.
- [8] A. V. Kuznetsov and B. S. Tsybakov, "Coding in a memory with defective cells," *Probl. Peredachi Inf.*, vol. 10, no. 2, pp. 52–60, Apr.–Jun. 1974.
- [9] E. Hwang, B. Narayanaswamy, R. Negi, and B. V. K. Vijaya Kumar, "Iterative cross-entropy encoding for memory systems with stuck-at errors," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Houston, TX, USA, Dec. 2011, pp. 1–5.
- [10] A. Jagmohan, L. A. Lastras-Montano, M. M. Franceschini, M. Sharma, and R. Cheek, "Coding for Multilevel Heterogeneous Memories," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Cape Town, South Africa, May 2010, pp. 1–6.
- [11] L. A. Lastras-Montano, A. Jagmohan, and M. M. Franceschini, "Algorithms for memories with stuck cells," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Austin, TX, USA, Jun. 2010, pp. 968–972.
- [12] A. V. Kuznetsov and A. J. H. Vinck, "On the general defective channel with informed encoder and capacities of some constrained memories," *IEEE Trans. Inf. Theory*, vol. 40, no. 6, pp. 1866–1871, Nov. 1994.
- [13] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge, U.K.: Cambridge University Press, 2011.
- [14] C. Heegard and A. El Gamal, "On the capacity of computer memory with defects," *IEEE Trans. Inf. Theory*, vol. 29, no. 5, pp. 731–739, Sep. 1983.
- [15] C. Heegard, "Partitioned linear block codes for computer memory with "stuck-at" defects," *IEEE Trans. Inf. Theory*, vol. 29, no. 6, pp. 831–842, Nov. 1983.
- [16] B. S. Tsybakov, "Additive group codes for defect correction," *Probl. Peredachi Inf.*, vol. 11, no. 1, pp. 111–113, Jan.–Mar. 1975.
- [17] Y. Kim and B. V. K. Vijaya Kumar, "Coding for memory with stuck-at defects," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Budapest, Hungary, Jun. 2013, pp. 4347–4352.
- [18] —, "Redundancy allocation of partitioned linear block codes," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Istanbul, Turkey, Jul. 2013, pp. 2374–2378.
- [19] A. Barg, *Lecture Notes on Error Correcting Codes*. [Online]. Available: <http://www.ece.umd.edu/~abarg/626/626-PartIII.pdf>
- [20] F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error-Correcting Codes*. Amsterdam, The Netherlands: North-Holland, 1977.
- [21] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge, U.K.: Cambridge University Press, 2004.

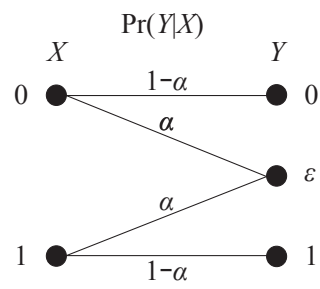


Fig. 1. Binary erasure channel (BEC).

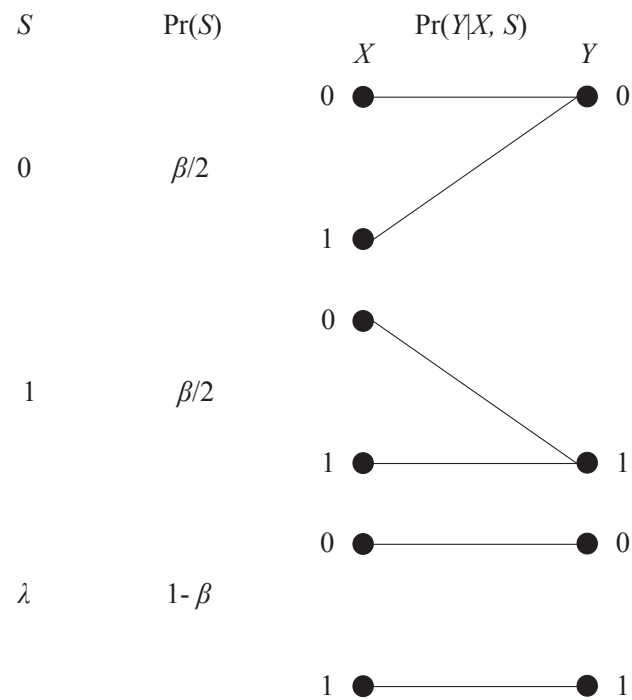


Fig. 2. Binary defect channel (BDC).

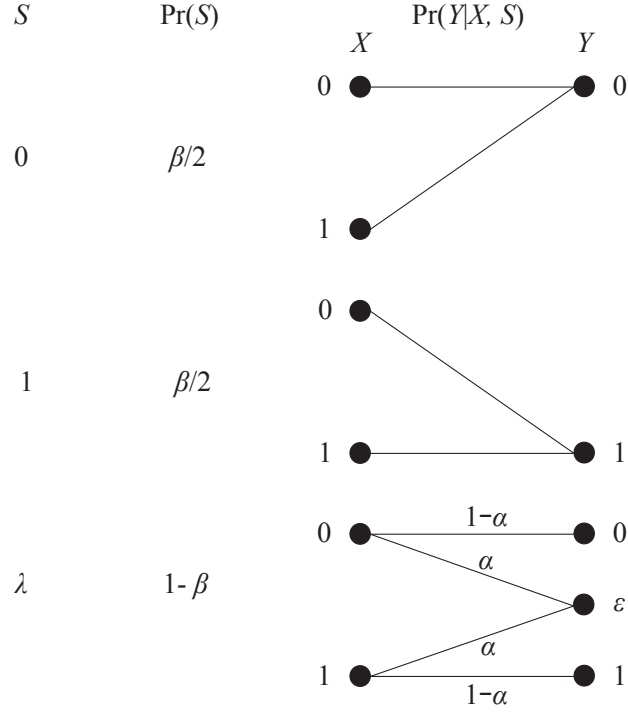
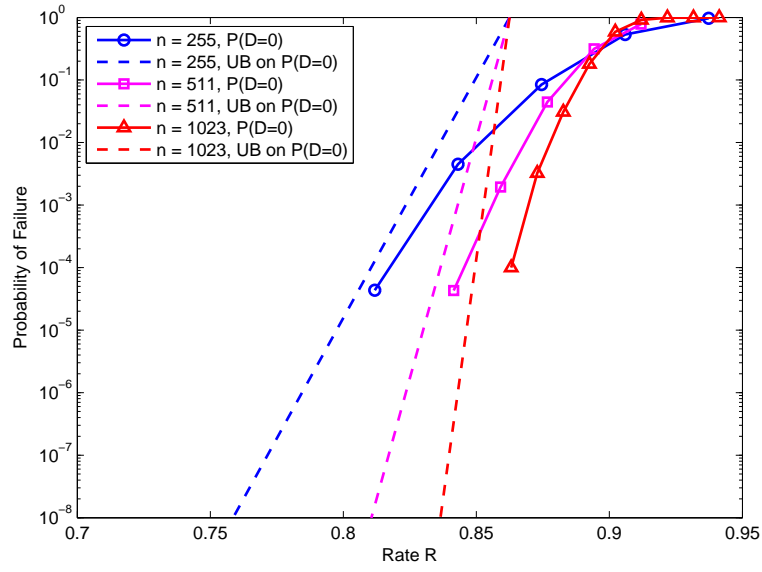


Fig. 3. Binary defect and erasure channel (BDEC).

Fig. 4. Probability of decoding failure, i.e., $P(D=0)$ for the BEC with $\alpha = 0.1$.

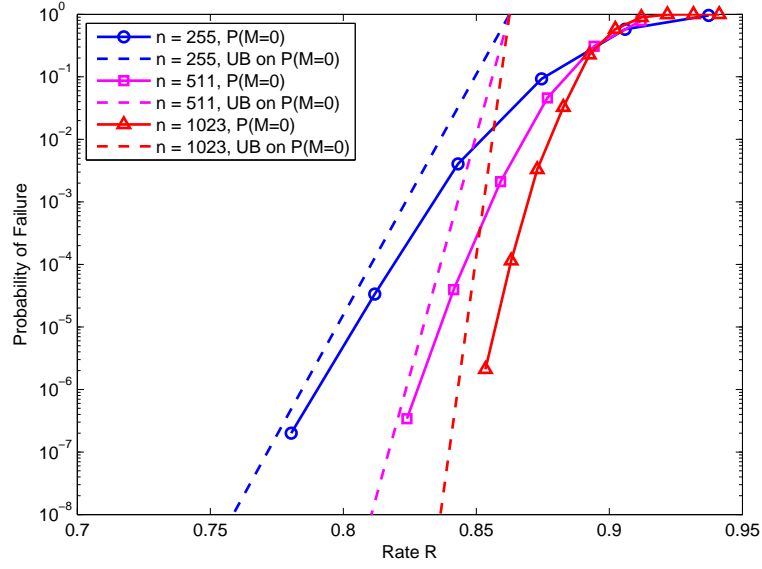


Fig. 5. Probability of masking failure, i.e., $P(M=0)$ for the BDC with $\beta = 0.1$.

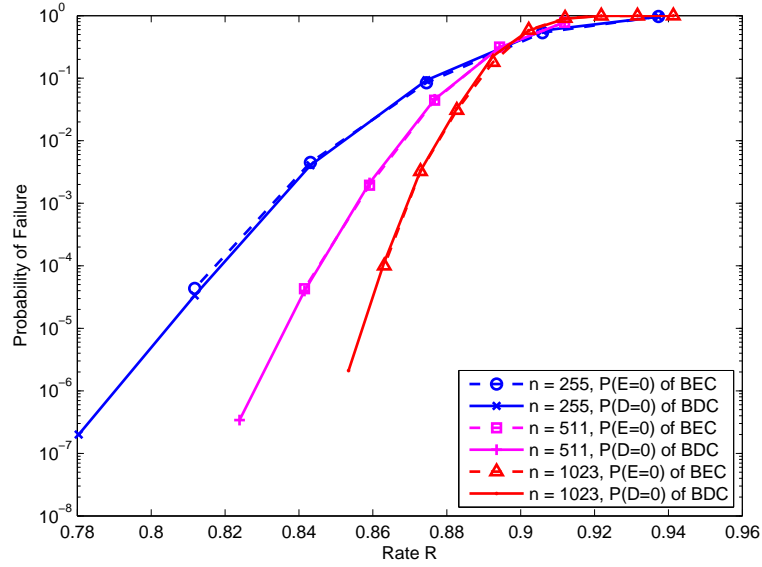


Fig. 6. Probability of failure, i.e., $P(D=0)$ for the BEC with $\alpha = 0.1$ and $P(M=0)$ for the BDC with $\beta = 0.1$.

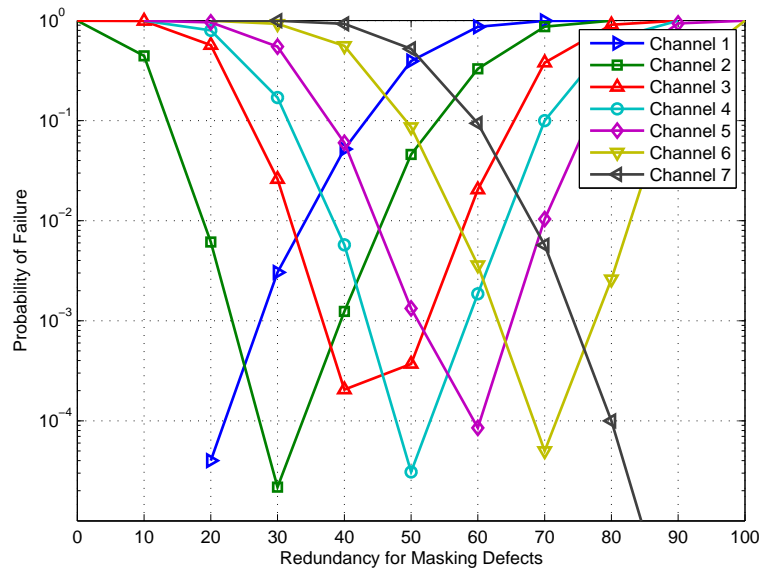


Fig. 7. Probability of failure for the channels in Table III.

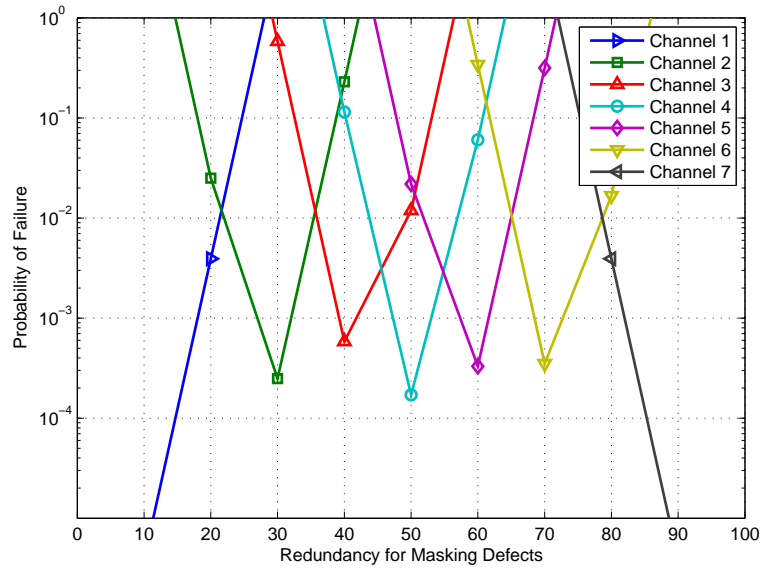


Fig. 8. Upper bound on probability of failure for the channels in Table III.